ENVIRONMENTAL RESEARCH

RESEARCH AND TECHNOLOGY BRANCH

Environment
Ontario
Jim Bradley, Minister

ISBN 0-7729-6907-8

DEVELOPMENT OF MULTIVARIATE ANALYSIS PROCEDURES FOR ONTARIO AIR QUALITY DATA

R. A. C. PROJECT NO. 311 PL
FINAL REPORT

Prepared for Environment Ontario by:

Philip K. Hopke
Institute For Environmental Studies
University of Illinois at
Urbana-Champaign

MARCH 1990

Copyright: Queen's Printer for Ontario, 1990
This publication may be reproduced for non-commercial purposes with appropriate attribution.

90-05626/

DISCLAIMER

This Report has been reviewed by the Research Advisory Committee of Environment Ontario and approved for Publication. Approval does not necessarily signify that the contents reflect the views and/or policies of Environment Ontario nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Table of Contents

	Page
Table of Contents	ii
List of Tables	iii
List of Figures	iv
Introduction	1
Objectives	1
Methodological Studies of Three-Mode Factor Analysis	2
Introduction	2
Review of the Unrotated Solution	3
Studies of Axis Rotation	7
Conclusion	10
Sources of Acidity in Wet and Dry Deposition	10
Introduction	10
Data Base	12
Data Analysis	14
Factor Analysis	14
Factor Analysis Applied to the Precipitation Chemical Data Sets	14
Factor Analysis Applied to the Combined Data Sets of Chemical and Back Trajectory Data	16
Potential Source Contribution Function	20
PSCF Based on Factor Scores	21
PSCF Based on Individual Species	28
Joint Probabilities	28
Conclusions	28
Local Scale Particle Source Apportionment	29
Introduction	29
Principle of Mass Conservation	30
Chemical Mass Balance	31
Multivariate Receptor Models	32
Target Transformation Factor Analysis	37
Source Apportionment in Hamilton	37
Fine Fraction Results	39
Summary	47
Acknowledgement	48
References	52

List of Tables

Table I. Factor loadings of the first mode. 5
Table II. Factor loadings of the second mode. 5
Table III. Factor loadings for the third mode. 6
Table IV. Unrotated core matrix. 6
Table V. Primary meteorological regimes and their relationship to factors and time periods. 6
Table VI. Varimax rotated core matrix. 8
Table VII. Average values and standard deviations for the precipitation composition variables. 14
Table VIII. Correlation coefficients between composition variables at station 3011. 15
Table IX. Varimax rotated factor loadings, variances, and communalities (composition data only). 16
Table X. Varimax rotated loadings for combined composition and trajectory endpoint data. 18
Table XI. Results of eigenvalue analysis and data reproduction tests for fine fraction data at Hamilton site 29025. 40
Table XII. Iterated vectors from target transformation analysis of Hamilton fine particle composition data. 41
Table XII (continued). Iterated vectors from target transformation analysis of Hamilton fine particle composition data. 42
Table XIII. Results of the regression analysis to obtain the scaling factors for the iterative TTFA profiles, Hamilton fine particle data. 43
Table XIV. Source profiles for fine particle Hamilton data (Values are as weight percent). 44
Table XV. Mass contributions of the various sources to the observed elemental concentrations. 46

List of Figures

Figure 1. Schematic outline of the simulated system. 3
Figure 2. Wind strengths and directions for the 12 simulation time periods. 4
Figure 3. The average contributions calculated from true values of corresponding sites and periods that belong to the specific zone and regime. 9
Figure 4. APIOS Event Wet/Dry Deposition Network Station Location Map.
a: stations 1011 and 1021; b: stations 1031 and 2011;
c: stations 3011 and 3021; d: stations 3031 and 3041;
e: stations 4011, 4021, 4031, and 4041; f: stations 6051, 6061, 6071, and 6081. 13
Figure 5. Subregion boundaries for eastern North America. 17
Figure 6. PSCF for factor 1 at station 3011. 22
Figure 7. PSCF for factor 2 at station 3011. 22
Figure 8. PSCF for factor 3 at station 3011. 23
Figure 9. PSCF based on pH at station 3011. 24
Figure 10. PSCF based on $SO_4^{=}$ at station 3011. 24
Figure 11. PSCF based on NO_3^{-} at station 3011. 25
Figure 12. PSCF based on Ca^{2+} at station 3011. 25
Figure 13. PSCF based on Na^{+} at station 3011. 26
Figure 14. Joint probability plot for stations 3011 and 4011. 26
Figure 15. Joint probability plot for stations 4011 and 1011. 27
Figure 16. Joint probability plot for stations 3011 and 1011. 27
Figure 17. Dendrogram of the iterated vectors from the TTFA analysis of the Hamilton fine particle data set. 43
Figure 18. Mass contribution for motor vehicles to the fine particle mass concentration in Hamilton. 49
Figure 19. Mass contribution for the Salt? source to the fine particle mass concentration in Hamilton. 49
Figure 20. Mass contribution for steel source to the fine particle mass concentration in Hamilton. 50
Figure 21. Mass contribution for flyash/soil source to the fine particle mass concentration in Hamilton. 50
Figure 22. Mass contribution of the regional sulfate aerosol to the fine particle mass concentration in Hamilton. 51
Figure 23. Mass contribution for the non-ferrous metal source to the fine particle mass concentration in Hamilton. 51

Introduction

The accumulation of analytical data characterizing environmental systems such as the atmosphere is only the first step in the accumulation of information needed for the development of management strategies to preserve or improve the quality of those systems. The Ministry of the Environment (MOE) has a large air quality monitoring program in place that produces substantial quantities of high quality data on both a local and regional scale. The development of improved automatic sampling and multiple species chemical analysis systems allow large multivariate data sets to be obtained. These data require careful multivariate statistical methods in order to extract the maximum amount of useful information. It is the purpose of this proposal to explore the utility of a variety of multivariate statistical methods to several distinctly different types of air quality data being obtained in Ontario and to develop efficient and effective data analysis procedures that will permit the Province to make maximum use of their on-going data gathering efforts. During the two years of support from the Ministry, we have begun to examine some of the multivariate methods that can be used to extract useful information regarding the origins of airborne pollutants and their transport in the environment from these data sets and to provide these methods in the form of computer programs to the Air Resources Branch for their use.

Objectives

The primary objective of this work is to determine if multivariate data analysis methods can extract evidence for long range transport of acidic materials influencing the precipitation chemistry of samples collected by the APIOS network. The secondary objective is the establishment of a set of data analysis procedures to be applied to these data as they are accumulated so that such information can be routinely extracted as part of the on-going data interpretation efforts. Finally, the ability of target transformation factor analysis for extraction of source information for urban scale aerosol will be explored and procedures will be established to incorporate this methodology in the techniques available to MOE for local air quality management of particulate emission sources. The initial studies completed at the University of Illinois are described in this report.

Methodological Studies of Three-Mode Factor Analysis

INTRODUCTION

Data collected for ambient air quality management generally can be organized as a three-dimensional data matrix with determined species, sampling sites, and sampling periods as the three dimensions. These data could be analyzed by a variety of methods in receptor modeling (Hopke, 1985) to investigate the sources and transport of the pollutants. However, all the existing methods in receptor modeling only analyze 1) the data for one sample, or 2) a set of samples collected at one site over periods or one period over sites or one species over a set of periods and sites. In other words, they use only one column (a vector) or one plane (a two-dimensional matrix) orientated in a certain direction in the complete data block (three-dimensional matrix). Since only a part of the data is used, the information obtained in these ways might be distorted and some information about the interaction among the dimensions might be lost. In order to analyze the data as a whole, Zeng and Hopke (1989) introduced three-mode factor analysis (denoted as TMFA hereafter) into receptor modeling. With TMFA, the spatial and temporal variations in the data can be investigated simultaneously.

TMFA was originally introduced by Tucker (1966). Since then, it has been primarily applied in the social sciences (Kroonenberg, 1983). The general principles of TMFA have been described by Tucker (1966) and Kroonenberg (1983). The basic idea is to extend the two-mode (conventional) factor analytical model to three-dimensional data. TMFA works on a three-dimensional data cube; each dimension corresponds to a class of variables, i.e. a "mode". The term "mode" is used to mean a "set of indices by which data might be classified" (Tucker, 1966). Through TMFA, the data cube is decomposed into three two-dimensional matrices (called factor loading matrices) and one three-dimensional matrix (called the core matrix). Similar to conventional factor analysis, most of the variation of measured variables is compressed into a few factors according to the covariance among the variables.

The application of three-mode analysis to aerosol receptor modeling has been described by Zeng and Hopke (1989). Three-dimensional observation data, x_{ijk}, can be decomposed by TMFA to yield the following model,

$$x_{ijk} = \sum_{m} \sum_{p} \sum_{q} a_{im} b_{jp} c_{kq} g_{mpq} \qquad (1)$$

Where a_{im}, b_{jp}, and c_{kq} are elements of factor loading matrices A, B, and C, respectively, and g_{mpq} is an element of the core matrix G. In this preliminary work, Zeng and Hopke (1989) used a simulated data set to study the applicability of TMFA in receptor modeling. The interpretable results have been obtained by introducing the concepts of "pollution zone" and "primary meteorological regime" in addition to "source". The corresponding relationships between the factors produced by TMFA and pollution sources, pollution zones and meteorological conditions were obtained by examining the factor loading matrices. Zeng and Hopke (1989) also described the difficulties in the interpretation of their unrotated factor solutions. They suggested that rotation was needed to improve the method. Several rotations, both orthogonal (such as varimax) and oblique (such as oblimax), were tested.

REVIEW OF THE UNROTATED SOLUTION

A data set was constructed based on a simulated airshed (Figure 1) using the Ontario acidic preciptiation network sampling sites as the receptor sites (R1-R8) and three classes of pollution sources. The first source type (S1) were local coal-fired power plants distributed as

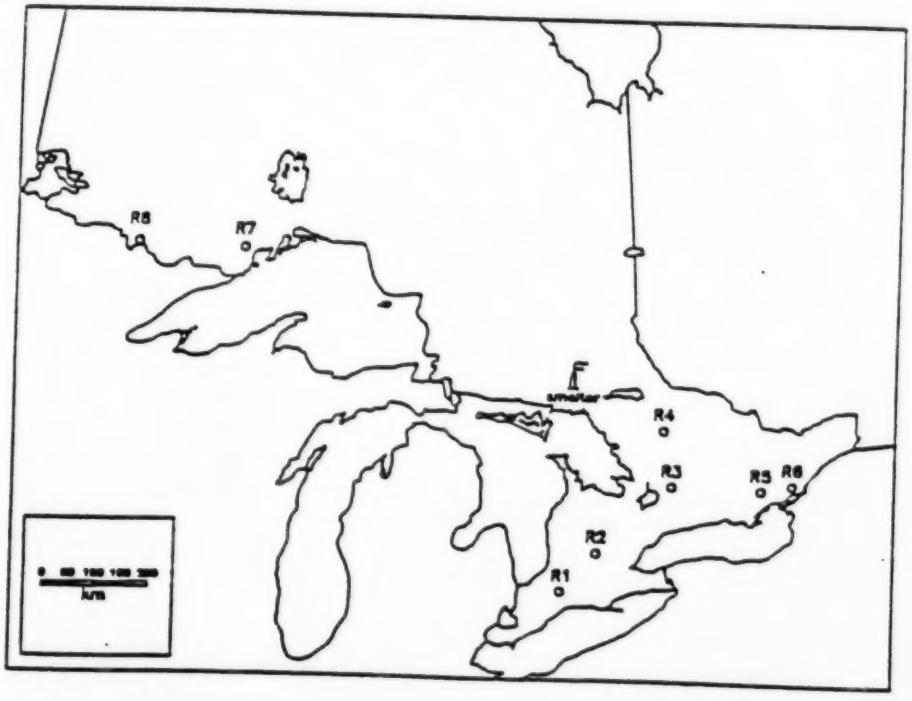

Figure 1. Schematic outline of the simulated system.

area sources around the receptors. The second source type (S2) was long-range transported power plant emissions from the midwestern U.S. The third source (S3) was a nickel smelter point source in Sudbury, Ontario. The samples were created for 12 periods that had different wind directions and wind strengths (Figure 2). For each period source contributions were estimated and the data set constructed. It was then analyzed by the TMFA program, TUCKALS3 (Kroonenberg and Brouwer, 1985). Modes 1, 2, and 3 (i, j, and k) correspond to chemical species, sampling site, and period, respectively. The number of retained factors was 3, 2, and 3, respectively. The fitted sum of squares was 95% of the total. The unrotated loading matrices A, B, C are given in Tables I, II, III, respectively, and core matrix is in Table IV.

The factor 1, mode 1, loadings had the same pattern as the chemical composition (i.e. source profile) of source 1 (S1). Factors 2 and 3 showed similar relationships with sources 3 and 2, respectively. In the mode 2 factor loadings (Table II), sites R7 and R8 showed the same behavior (both factor 1 and 2 have positive loadings on them), and other sites show a different pattern (positive loading for factor 1, negative for factor 2), dividing the sites into two groups. One group (R7 and R8) is in a cleaner area (zone 2). The other sites (R1-R6) is a polluted area (zone 1). Table III contains ${}_tC_q$. The 12 periods can be classified into 3 groups

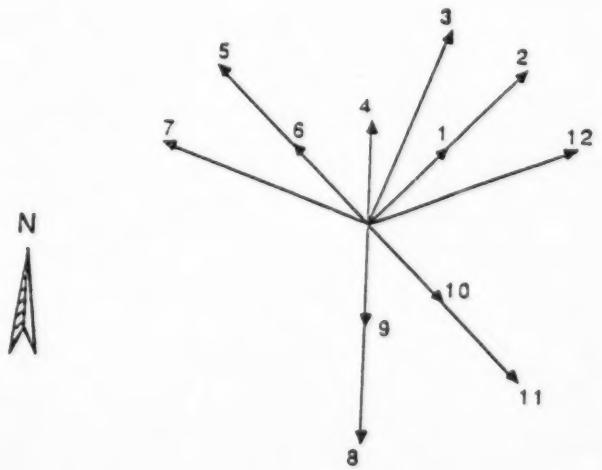

Figure 2. Wind strengths and directions for the 12 simulation time periods.

Table I. Factor loadings of the first mode.

i	Element	Unrotated m=1 (S1)	Unrotated m=2 (S3)	Unrotated m=3 (S2)	Varimax rotated m'=1 (S1)	Varimax rotated m'=2 (S3)	Varimax rotated m'=3 (S2)
1	Na	-0.046	0.047	0.009	-0.053	0.038	0.012
2	Al	0.051	0.024	-0.050	0.059	0.044	-0.014
3	Si	0.871	-0.112	-0.274	0.920	0.024	0.019
4	S	0.333	0.386	0.745	0.028	-0.004	0.902
5	Cl	-0.004	0.096	0.062	-0.037	0.057	0.092
6	K	-0.041	0.059	0.005	-0.049	0.050	0.016
7	Ca	0.051	0.048	-0.002	0.041	0.044	0.036
8	Ti	-0.103	0.052	-0.007	-0.102	0.050	-0.018
9	V	-0.115	0.048	0.010	-0.118	0.039	-0.010
10	Cr	-0.119	0.049	-0.002	-0.119	0.044	-0.021
11	Mn	-0.090	0.066	0.071	-0.117	0.026	0.057
12	Fe	0.043	0.183	-0.054	0.029	0.187	0.048
13	Ni	0.015	0.854	-0.453	0.020	0.966	-0.003
14	Cu	-0.102	0.058	0.021	-0.111	0.042	0.008
15	Zn	0.054	0.153	0.380	-0.088	-0.039	0.402
16	As	-0.109	0.053	0.028	-0.119	0.035	0.009
17	Se	-0.120	0.047	0.004	-0.121	0.040	-0.017
18	Br	-0.118	0.050	0.005	-0.110	0.043	-0.014
19	Sb	-0.121	0.047	0.002	-0.122	0.041	-0.019
20	Pb	-0.069	0.100	0.067	-0.100	-0.058	0.075

Table II. Factor loadings of the second mode.

j	Site	Unrotated p=1 (zone 1)	Unrotated p=2 (zone 2)	Varimax rotated p'=1 (zone 1)	Varimax rotated p'=2 (zone 2)
1	R1	0.459	-0.095	0.468	0.023
2	R2	0.455	-0.100	0.465	0.017
3	R3	0.405	-0.133	0.425	-0.028
4	R4	0.382	-0.126	0.401	-0.027
5	R5	0.326	-0.082	0.336	0.003
6	R6	0.322	-0.078	0.331	0.005
7	R7	0.175	0.691	-0.004	0.713
8	R8	0.181	0.676	0.006	0.700

Table III. Factor loadings for the third mode.

k (period)	Unrotated q=1 (regime ?)	Unrotated q=2 (regime ?)	Unrotated q=3 (regime ?)	Varimax rotated q*=1 (regime 1)	Varimax rotated q*=2 (regime 2)	Varimax rotated q*=3 (regime 3)
1	0.301	0.071	-0.207	0.372	-0.014	0.013
2	0.291	0.153	0.452	-0.006	0.016	0.559
3	0.289	0.142	0.482	-0.027	0.004	0.579
4	0.303	0.047	-0.207	0.368	-0.037	0.008
5	0.263	-0.542	0.039	0.064	-0.599	0.038
6	0.308	0.014	-0.147	0.329	-0.075	0.051
7	0.208	-0.762	0.063	-0.040	-0.790	-0.028
8	0.301	0.092	-0.131	0.331	0.000	0.079
9	0.303	0.099	-0.291	0.429	0.019	-0.046
10	0.284	0.166	-0.380	0.481	0.095	-0.112
11	0.307	0.088	-0.062	0.295	-0.011	0.137
12	0.291	0.149	0.449	-0.005	0.012	0.555

Table IV. Unrotated core matrix.

	Regime 1 (q=1) Zone 1 (p=1)	Regime 1 (q=1) Zone 2 (p=2)	Regime 2 (q=2) Zone 1 (p=1)	Regime 2 (q=2) Zone 2 (p=2)	Regime 3 (q=3) Zone 1 (p=1)	Regime 3 (q=3) Zone 2 (p=2)
Source 1 (m=1)	40.091	0.299	-0.198	-4.234	-0.062	-0.685
Source 2 (m=3)	0.036	-1.591	0.158	0.431	8.164	-2.010
Source 3 (m=2)	-1.136	-0.886	-0.218	-10.862	0.483	0.

Table V. Primary meteorological regimes and their relationship to factors and time periods.

	Primary Meteorological Regime 1		2	3
Corresponding unrotated factor	Factor 1 ? (q=1) (q=2)		Factor 2 ? (q=3)	Factor 3 ?
Corresponding varimax rotated factor	Factor 1 (q*=1)(q*=2)		Factor 2 (q*=3)	Factor 3
Wind direction	All direction	Northern OR	SE	SW
Wind speed	Weak	Strong	Strong	Strong
Corresponding periods	1, 4, 6, 9, 10	8, 11	5, 7	2, 3, 12

corresponding to 3 primary meteorological regimes (Table V). The third regime is strong SW wind where the smelter effect on all sites is small. The local source contribution is moderate. The wind brought large quantities of long range pollutants to zone 1. The second regime was strong SE wind where the contribution of local sources to all sites is relatively small, but the wind blew pollutants from the distant coal sources and the smelter to zone 2. In regime 1,weak wind, long-range transport was not important. The major contributor to all sites are the local sources. Since upper Ontario was assumed to be a clean area, northern winds and high wind speed (period 8, 11) blew local pollutants away and no long-range transported material arrived at the receptors. The pollutants were less than in low wind conditions. Therefore, periods 8 and 11 had similar loadings to those of the weak wind condition periods (1, 4, 6, 9, and 10).

As observed in the core matrix, the variation at zone 1 (p = 1) was primarily accounted for by local sources (m = 1) [g_{111} = 40.091] under the meteorological regime 1 (q = 1, weak or northern wind). For a strong SE wind condition (q = 2), zone 2 (p = 2) was affected by the Ni-smelter (m = 2), and for the strong SW wind condition (q = 3), distant sources (m = 3) pollute zone 1.

STUDIES OF AXIS ROTATION

The unrotated solutions have shown a limited relationship with the simulated "physical model", but there were several points that were difficult to interpret. Factor 1 of mode 3 had very similar loadings for every period (Table III). The first factor usually carries the highest amount of information compared with the others, but here we cannot draw any useful conclusions from this factor. In the core matrix, the values of the matrix elements seem not to be well partitioned. Note that the g^2_{mpq} values are the variance explained by the combination of factors m, p, and q in the three different modes. In Table IV, g_{111} = 40.09, and the other g_{mpq} (q = 1) are negligible. Thus, for the weak or northern wind conditions (q = 1), the local sources (m = 1) totally dominate such that the other sources are totally negligible. However, this result is not consistent with the simulated "physical model" in which the distant coal sources (S2) also make a significant contribution to most receptors. There are similar problems for the other conditions (q = 2, 3). In addition, the corresponding relationships between the source profiles and the factors in mode 1 are not very good. These results suggest the need for a

rotation of the abstract factor matrices to achieve a kind of "simple structure" as in conventional factor analysis (Hopke, 1985).

Rotation can be individually performed on each of the three loading matrices as in conventional factor analysis. After successful rotations are achieved, the inverse transformation can be applied to the core matrix. Varimax (Kaiser, 1958) rotation was employed both with or without row normalization (factor loadings of each variable are normalized by the corresponding communality). The similarity of the source profiles and the loadings of mode 1 can be measured by correlation coefficients. In general, the correlations are increased after rotation. The varimax without normalization gave the best results. Therefore, it is selected and referred to as varimax for simplicity.

The varimax rotated loadings are listed with unrotated loadings in Table I. The varimax rotated loadings are listed in Table II and III for mode 2 and 3, respectively. The basic pattern of the loadings on mode 2 remains the same after rotation. However, the result for mode 3 is significantly improved by the rotation. The loadings of factor 1 in mode 3 (see Table III) are no longer uniform. Only the values for periods 1, 4, 6, 8, 9, 10, and 11 (i.e. the regime 1) are significant. These results are consistent with the actual meteorological regimes if we assign rotated factor 1 to meteorological regime 1, factor 2 to regime 2, and factor 3 to regime 3.

The rotated core matrix presented in Table VI provides a more reasonable partition of the system variance. This point can be seen by comparing the core values with the true source contributions (Figure 3). The average true contributions are calculated from corresponding sites and periods that belong to the specific zone and regime. Under weak wind or strong northern wind condition (meteorological regime 1, $q^*=1$), the local sources (S1, $m^*=1$) dominate the

Table VI. Varimax rotated core matrix.

	Regime 1 ($q^*=1$)		Regime 2 ($q^*=2$)		Regime 3 ($q^*=3$)	
	Zone 1 ($p^*=1$)	Zone 2 ($p^*=2$)	Zone 1 ($p^*=1$)	Zone 2 ($p^*=2$)	Zone 1 ($p^*=1$)	Zone 2 ($p^*=2$)
Source 1 ($m^*=1$)	29.497	7.745	-11.196	-5.580	18.317	4.694
Source 2 ($m^*=3$)	6.168	-0.121	-3.473	-5.881	13.445	-0.109
Source 3 ($m^*=2$)	1.818	-2.906	2.714	-9.013	-3.009	-1.000

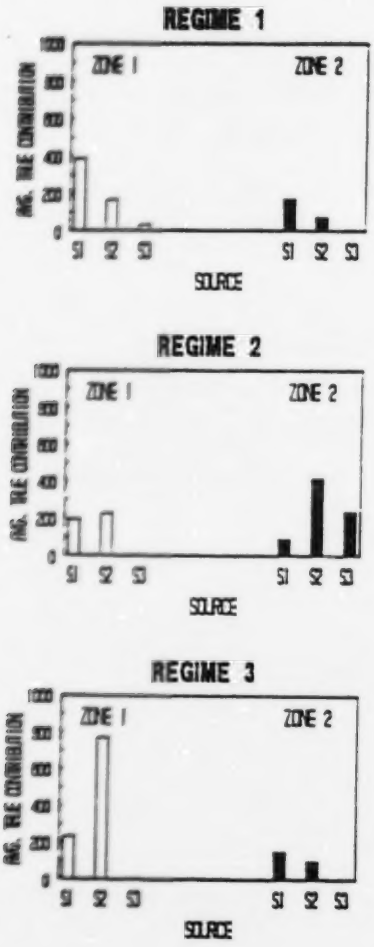

Figure 3. The average contributions calculated from true values of corresponding sites and periods that belong to the specific zone and regime.

contributions to both zone 1 and 2 ($p^*=1$ and 2) [$g^2_{\alpha pq}$ are $(29.49)^2$ and $(7.74)^2$, respectively]. Meanwhile, it can be seen that the regional coal source (S2, $m^*=3$) is also significant in zone 1 [$(6.17)^2$]. Other situations are negligible in this regime. With a strong southeast wind (regime 2, $q^*=2$), there are also some local source contributions in both zones [$(-11.20)^2$ and $(-5.58)^2$], but less than in regime 1 (see Figure 3). Relatively large contributions of Ni-smelter (S3, $m^*=2$) and regional coal (S2, $m^*=3$) to zone 2 are found in this situation [$(-9.01)^2$ and $(-5.88)^2$, respectively]. For strong southwest wind, there is a very significant contribution from regional coal (S2, $m^*=3$) to zone 1 [$(13.44)^2$] in addition to the local source contribution as under the

other conditions. This interpretation is much more reasonable than the one that could be made for the unrotated core matrix (Table IV). The rotated core matrix has qualitatively achieved agreement with the simulated "physical system" in terms of average source contribution.

CONCLUSION

A substantial improvement was achieved by rotating the factors produced through TMFA. The varimax (without row normalization) was found to be the most useful rotation for the loading matrix of each of the three modes. The inverse transformation was then performed on the core matrix. The rotated loading and core matrix were more easily interpreted and tended to be unique. The TMFA model tended to agree with the simulated physical system after the rotation. Thus, the initial development of TMFA has been completed and we are now examining data sets to find suitable candidates for analysis. The analysis of real data is the next step in the development of TMFA to being a practical and useful tool in receptor modeling. The results presented here have been accepted for publication (Zeng and Hopke, 1990).

Sources of Acidity in Wet and Dry Deposition

INTRODUCTION

Acidic precipitation is a serious regional environmental problem in eastern North America. In order to understand the origins of this phenomenon and then develop corresponding strategies to decrease the acidity of the precipitation, it is necessary to understand the chemical characteristics of the sources of the acid precipitation and their geographical positions. Although it is known that sulfate and nitrate transformed from SO_2 and NO_x in the atmosphere are major acid species and the distribution of emissions of SO_2 and NO_x over North America is available, it may not be necessarily true that high SO_2 or NO_x emission rate areas are the dominate contributing source areas to the acidity of precipitation at a specific receptor in North America. Therefore, a study is necessary for a receptor area to offer evidence to show where the sources are located.

The formation and transport of acidic precipitation is a complicated process. If a dispersion model is used, comprehensive studies are needed that include a survey of emissions over a large scale, collection of meteorological data, field tests to determine the dispersion

parameters, and mechanistic studies on the chemical transformation and scavenging of gases and particles. In addition, the dispersion model can only handle the sources that are known to contribute to acid precipitation, essentially the identified SO_2 and NO_x emission sources. It will not identify the possibilities of other sources.

Receptor models have been used in regional pollution aerosol studies (Rahn and Lowenthal, 1984, 1985; Lowenthal and Rahn, 1988; Lowenthal *et al.*, 1988), especially in the studies of aerosol sulfate. Basically, these studies belong to tracer methods or chemical mass balance type of models in receptor modeling (Hopke, 1985). In principle, receptor models can use the chemical measurements of the precipitation samples collected from monitoring sites (receptors) as the data base, identify the possible sources in terms of their chemical nature, and estimate the importance or contribution of these sources. In these studies, regional source profiles have been developed to identify the general area from which the acidic material has come.

However, the normal receptor models can not determine the specific location of sources. To overcome this shortcoming, meteorological information needs to be incorporated into the receptor model. Keeler and Samson (1987) presented a method to test the results of receptor modeling for long range transport of airborne particles using meteorological data. Malm *et al.* (1986) proposed another approach in their study of the sources of particles at the Grand Canyon, Arizona. In this approach, an air parcel trajectory model (Bresch *et al.*, 1984) was used to determine the transport path to the receptor site. For each sample, the back trajectories ending at the receptor site were calculated. The region surrounding the receptor site was divided into subregions. A new variable was defined as the number of endpoints of segments of trajectories (the output from the trajectory model) shown in each of the defined subregions. This variable reflected the residence time of the air parcel in the specific subregion, and was used in factor analysis (Hopke, 1985) with the chemical data. The factor loadings showed the relations between the chemical characteristics of the sources and residence time of the air parcels in the sources areas.

In order to study the sources of acidity in wet deposition, the precipitation chemistry data sets, will be investigated with the emphasis on the chemical characteristics and geographic location of the sources. The approaches described by Malm *et al.* were used to study the sources of precipitation constituents in Ontario.

DATA BASE

The Acidic Precipitation in Ontario Study (APIOS) Event Wet/Dry Deposition Network has run since 1980 (Chan *et al.*, 1985). This network consists of 16 stations over lower Ontario (Figure 4). The stations 1 to 13 are aligned along the prevailing wind direction for the southern part of Ontario and these areas have significant levels of precipitation acidity based on historical data. Measurements are made on the wet deposition samples that are collected daily. The variables measured are sample volume in ml (VOLUME), conductivity at 25°C in μmho/cm (COND25), field pH (FWPH), pH (PH), total acidity in mg $CaCO_3$/l (ACDT), sulfate in mg SO_4^{2-}/l (SSO4UR), nitrate in mg N/l (NNO3UR), calcium in mg Ca^{2+}/l (CAUR), chloride in mg Cl^-/l (CLIDUR), magnesium in mg Mg^+/l (MGUR), potassium in mg K^+/l (KKUR), sodium in mg Na^+/l (NAUR), ammonium in mg N/l (NNHTUR), and total acidity in μg H/l (ACDG).

In this work, only three stations are considered: stations number 1011, 3011 and 4011 near London, Dorset and Kingston area, respectively. The investigation has been more concentrated on station 3011, Dorset. Data are available from 1980 to 1986. However, only the data for 1984-86 have been used because they are more complete and have been subjected to better quality control. Since more than half of data are missing for FWPH and ACDG, these two variables have been eliminated from the analysis. ACDT also has a considerable number of missing values. In an initial analysis, it was found that ACDT always contributes an unique factor in factor analysis. Thus, it does not provide useful information in this kind of analysis, and was also excluded from data sets used in this analysis. After eliminating the samples with identified data problems (such as unreliable data), the number of samples analyzed in this study are 312, 235, and 222 for station 3011, 4011, and 1011, respectively.

Air parcel back trajectories (BT) ending at each station have been calculated by Ontario Ministry of the Environment (MOE) (Olson *et al.*, 1978). The BT data are provided in the form of a list of time intervals and coordinates of the trajectory segment endpoints for each trajectory. Trajectories using surface level data are calculated each day at 0:00, 6:00, 12:00, and 18:00 using 3 hour time intervals for 48 hours backward in time. Trajectories at 850 mb level are also provided, but they are not suitable for this work because the time interval is 12 hours. Thus each trajectory segment spans too large a geographical region and the spatial resolution is

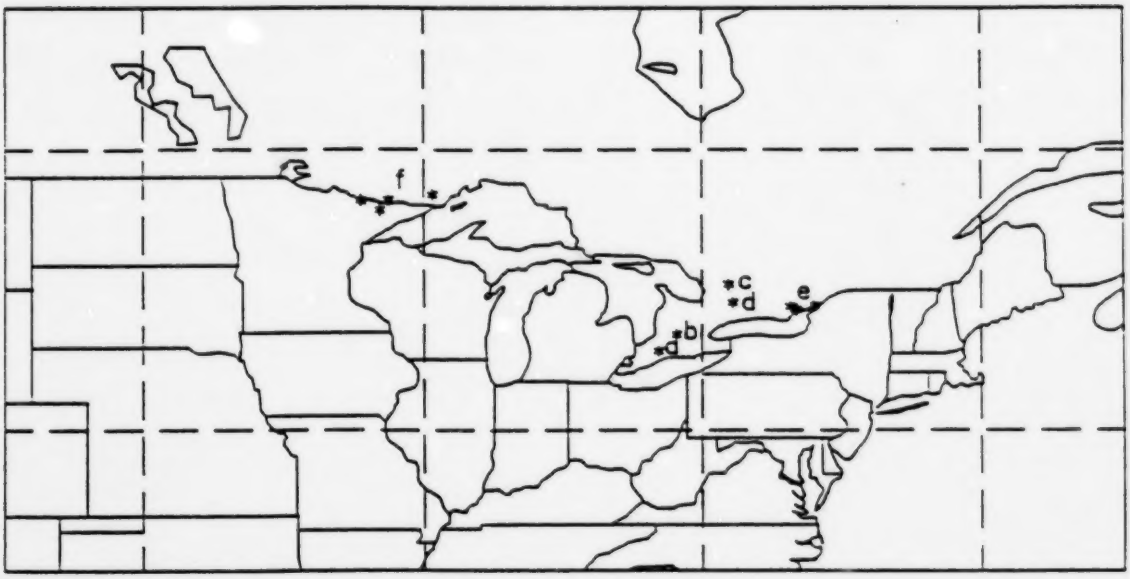

Figure 4. APIOS Event Wet/Dry Deposition Network Station Location Map. a: stations 1011 and 1021; b: stations 1031 and 2011; c: stations 3011 and 3021; d: stations 3031 and 3041; e: stations 4011, 4021, 4031, and 4041; f: stations 6051, 6061, 6071, and 6081.

insufficient to be useful in source identification.

DATA ANALYSIS

Table VII presents average value and standard deviation for each variable of the chemical data set at each station. The pH values for three stations (4.36, 4.28, and 4.35, respectively) are much lower than 5.6, the "natural" acidity of rainwater in the form of pH. From the correlation coefficient matrix given in Table VIII (only the data for station 3011 are listed; very similar relationships exist for other two stations), it can be seen that the correlation coefficients among COND25, pH, SO_4^{2-} and NO_3^- and between Ca^{2+} and Mg^{2+} are very high. The coefficients of SO_4^{2-} and NH_4^+, NO_3^- and NH_4^+, Cl^- and Na^+ are also high. To reveal the relations of variables and sources, factor analysis is used.

FACTOR ANALYSIS

Factor Analysis Applied on the Precipitation Chemical Data Sets

The data sets are reconstructed by multiplying every variable by VOLUME so that the variables are converted from concentrations to total amount (actually, converted to flux). The purpose of this transformation is to eliminate the dilution effect caused by the differences in precipitation volume. At the same time, the variable VOLUME is also excluded from further

Table VII. Average values and standard deviations for the precipitation composition variables.

Variable	Station 3011		Station 4011		Station 1011	
	Avg.	Std. Dev.	Avg.	Std. Dev.	Avg.	Std. Dev.
VOLUME (ml)	513.6	564.8	567.6	638.2	632.5	638.6
COND25(μmho/cm)	30.3	19.5	37.9	24.6	34.8	17.7
pH	4.356	0.307	4.279	0.433	4.352	0.408
SO_4^{2-} (mg/l)	2.45	2.03	3.07	2.48	3.49	1.96
NO_3^- (mg N/l)	0.565	0.404	0.712	0.537	0.646	0.392
Ca^{2+} (mg/l)	0.208	0.238	0.264	0.308	0.525	0.563
Cl^- (mg/l)	0.148	0.119	0.183	0.166	0.225	0.154
Mg^{2+} (mg/l)	0.033	0.040	0.035	0.043	0.066	0.071
K^+ (mg/l)	0.033	0.032	0.026	0.028	0.060	0.053
Na^+ (mg/l)	0.051	0.050	0.059	0.065	0.081	0.074
NH_4^+ (mg N/l)	0.297	0.276	0.305	0.276	0.401	0.279

Table VIII. Correlation coefficients between composition variables at station 3011.

	COND25	pH	SO_4^{2-}	NO_3^-	Ca^{2+}	Cl^-	Mg^{2+}	K^+	Na^+	NH_4^+
COND25	1.000									
pH	.981	1.000								
SO_4^{2-}	.962	.932	1.000							
NO_3^-	.913	.879	.825	1.000						
Ca^{2+}	.628	.541	.700	.660	1.000					
Cl^-	.797	.750	.733	.811	.608	1.000				
Mg^{2+}	.625	.534	.699	.637	.940	.605	1.000			
K^+	.652	.614	.651	.606	.499	.604	.516	1.000		
Na^+	.524	.466	.490	.538	.482	.759	.503	.559	1.000	
NH_4^+	.814	.733	.860	.812	.695	.687	.697	.595	.455	1.000

analysis. The number of variables in the final chemical data sets is ten. In addition, pH is transformed to $[H^+]$ in mMole/l, then to mMole by multiplying VOLUME. R-mode factor analysis with varimax rotation (Hopke, 1985) is performed on the chemical data sets for the three sites individually. The rotated factor loadings, communalities, and the explained variance for the three data sets are all listed in Table IX. From Table IX, it can be seen that about 90% of the variance in the system can be accounted for by taking three factors, and the communalities are very high for all of the variables except K^+. It has a somewhat lower value (0.6 to 0.7). Thus, three principal factors (three major sources) are indicated for this system.

The results are very similar among these three stations. The analyses for the three sites indicate three factors. Factor 1 always has large loadings on COND25, H^+, SO_4^{2-}, NO_3^- and medium loadings on NH_4^+ and Cl^-. Nearly 45% of variance in the system can be explained by it. It is the source responsible for the acidity of the precipitation because it relates H^+, SO_4^{2-} and NO_3^-.

In addition to this acid factor, there are two factors contributing to the precipitation, but not affecting the acidity of the precipitation. These two factors make up about 40-50% of the variance in the system. Factor 2 has large loadings on Ca^{2+} and Mg^{2+}, and medium loading on NH_4^+. A reasonable hypothesis is that particles containing Ca and Mg such as fine soil particles or other Ca and Mg enriched particles are incorporated into the precipitation. Ca^{2+} and Mg^{2+} ions then can be leached into rainwater. Other crustal elements such as Fe would have

Table IX. Varimax rotated factor loadings, variances, and communalities (composition data only).

Station	Factor 1			Factor 2			Factor 3			Communalities		
	3011	4011	1011	3011	4011	1011	3011	4011	1011	3011	4011	1011
COND25	.897	.941	.933	.288	.284	.201	.307	.120	.277	.982	.981	.987
H^+	.924	.945	.948	.181	.227	.056	.265	.096	.250	.956	.955	.965
SO_4^{2-}	.844	.902	.868	.417	.358	.363	.238	.076	.269	.943	.948	.958
NO_3^-	.799	.811	.842	.343	.327	.372	.345	.284	.278	.875	.845	.925
Ca^{2+}	.325	.334	.247	.892	.848	.888	.243	.041	.115	.960	.832	.862
Cl^-	.572	.530	.585	.284	.439	.244	.670	.578	.715	.857	.809	.913
Mg^{2+}	.308	.211	.244	.892	.926	.840	.268	.070	.275	.963	.908	.841
K^+	.493	.394	.341	.229	.630	.178	.569	.274	.738	.619	.627	.693
Na^+	.179	.091	.139	.231	.055	.157	.920	.965	.916	.932	.942	.883
NH_4^+	.713	.539	.792	.512	.446	.449	.209	.273	.232	.815	.564	.883
	Total Var. Explained											
Variance	4.32	4.14	4.46	2.45	2.74	2.33	2.13	1.53	2.13	8.90	8.41	8.92

provided additional information on this source, but unfortunately were not measured. Factor 3, with a large loading on Na and a medium loading on Cl, is considered to be sea salt particles that have been incorporated into the precipitation.

In summary, the factor analysis on chemical data suggest that there are three kinds of principal sources contributing the precipitation: acidic precursor sources, Ca and Mg sources, and Na and Cl sources. Special attention should be given to acid sources because they are responsible for the precipitation acidity. Now the question is where these sources come from in terms of geographical location. To investigate the location of the sources, the trajectory information is introduced into the analysis.

Factor Analysis Applied to the Combined Data Sets of Chemical and Back Trajectory Data

The trajectory information is in form of geographic distribution of back trajectory segment endpoints. Fourteen source subregions have been selected within the 2-day trajectory covered region (Figure 5). The setting of the subregion boundaries includes the consideration

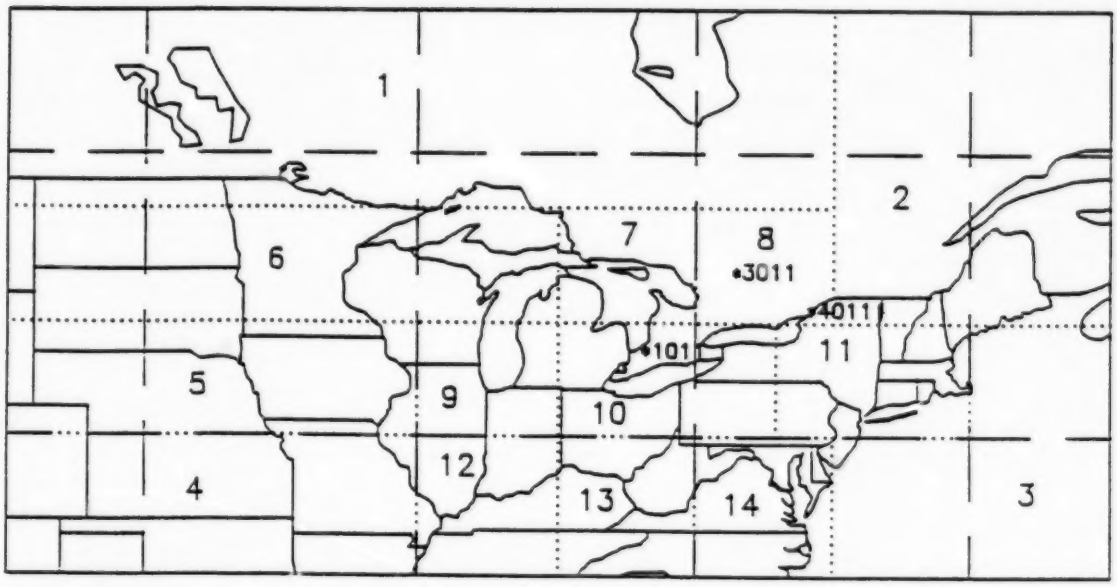

Figure 5. Subregion boundaries for eastern North America.

of separating potentially interesting source areas and subregion size. A subregion should not be too big, thus losing spatial resolution, nor too small so that there are too few endpoints in the subregion. Using these fourteen subregions, fourteen new variables are defined as the number of the segment endpoints in each subregion. The number of segment endpoints represents the number of 3 hour time intervals that the parcel remains in the subregion. Thus, the number of segment endpoints provides a measure of residence time in the subregion.

The factor analysis is applied to the combined data sets with 24 variables (10 chemical variables and 14 BT endpoint variables). Because of the additional spatial variables, additional factors (typically, a total of 10 to 11 factors) are needed to adequately reproduce the data. Additional factors are included until the chemical data are adequately reproduced. For all cases, only those factors that have loadings for the chemical variables similar to those given in Table IX are presented in Table X. A number of the factors only have significant loadings for the

endpoint variables. These factors reflect only the displacement of air mass from one subregion to another (Malm *et al.*, 1986). There are then two classes of factors. Only those factors that have significant loadings for the chemical variables are listed in Table X. The factors which associated only with endpoint variables are not included in Table X. The loadings in the range of 0.2 to 0.9 are considered to be weak (but significant) or moderate (Malm *et al.*, 1986).

In an ideal situation, if the trajectories only pass through specific pollutant source areas, i.e. the number of endpoints in these areas is large, the air mass will then carry significant amounts of pollutants from these sources to the receptor. The concentration of corresponding pollutants in the collected precipitation will be high. Alternatively, if trajectories only pass clean

Table X. Varimax rotated loadings for combined composition and trajectory endpoint data.

	Factor 1			Factor 2			Factor 3		
Station No.	3011	4011	1011	3011	4011	1011	3011	4011	1011
COND25	.973	.970	.970	.061	.108	.052	.015	.005	.143
H+	.954	.958	.955	-.035	.042	-.081	-.023	.000	.110
SO_4^{2-}	.928	.937	.935	.192	.184	.197	-.012	-.033	.148
NO_3^-	.918	.884	.911	.140	.186	.234	.036	.151	.172
Ca^{2+}	.587	.488	.401	.764	.733	.723	.077	-.006	.134
Cl^-	.822	.697	.702	.129	.328	.138	.341	.433	.624
Mg^{2+}	.582	.398	.417	.763	.825	.742	.083	.029	.261
K^+	.668	.541	.459	.078	.574	.128	.344	.160	.646
Na^+	.541	.233	.284	.168	.086	.119	.658	.859	.861
NH_4^+	.833	.638	.877	.291	.382	.306	-.025	.111	.130
Subregion 1	-.155	-.087	-.149	.029	-.109	.008	-.050	.215	-.028
Subregion 2	-.154	-.177	-.098	.030	.117	.004	-.051	.003	-.070
Subregion 3	-.096	-.045	-.003	.019	-.096	.095	-.004	.286	-.152
Subregion 4	-.027	-.056	-.032	-.016	.003	-.114	.153	-.113	.104
Subregion 5	-.047	-.023	-.083	-.100	-.098	-.195	-.139	.073	-.026
Subregion 6	-.157	-.063	-.152	-.112	-.100	.062	.041	.119	-.023
Subregion 7	-.103	-.077	-.084	.065	.162	-.118	-.069	-.133	-.155
Subregion 8	.010	.001	-.079	-.038	.032	-.244	-.116	-.336	-.084
Subregion 9	-.099	-.065	-.091	.105	.105	.349	-.131	.164	-.183
Subregion 10	.286	.270	.065	.214	.041	.104	-.153	-.183	-.043
Subregion 11	.031	.127	-.079	-.124	-.248	.053	-.025	-.078	.132
Subregion 12	.194	-.096	.194	.042	.117	.364	.031	.010	-.051
Subregion 13	.359	.209	.105	-.010	-.094	-.143	-.254	.237	.153
Subregion 14	.144	.006	.055	-.035	.015	-.322	.489	.091	.099

areas, the concentration of the pollutants will be low. Therefore, the number of endpoints in the areas (endpoint variables) and concentrations of corresponding pollutants will strongly covary. In the factor analysis, these endpoint variables and the chemical species (or pollutants) will show high loadings in the same factor. However, in actual cases, some trajectories pass both source and clean areas. This mixture of trajectories for a given precipitation event weakens the covariance so that the factor loadings on the endpoint variables are usually not very high. In addition, precipitation during transport from the source region to the receptor site will reduce the measured concentrations and weaken the covariance between the spatial and chemical variables. Thus, lower loading values can be expected compared with the results of Malm *et al.* for particulate composition data.

In Table X, according to the loadings on the chemical variables, factors 1, 2, 3 represent acid sources, Ca and Mg sources, and Na and Cl sources, respectively, as in the previous section. For the first factor at station 3011, subregions 13 and 10 show loadings of 0.359 and 0.286. They are significant values for endpoint variables. This result indicates that there are acid sources in these subregions, i.e. the Ohio River Valley region. Subregions 12 and 14, with the loadings of 0.194 and 0.144, are considered to be weak source regions. The remaining loadings are very small or negative. Negative loadings mean that either the trajectories passed these subregions and the concentrations of acid species at the receptor are low, or when the concentrations of acid species at the receptor are high, there is little probability that the trajectories passed these subregions. Thus, these subregions are clean or relatively clean regions with respect to acid sources.

Similarly, at station 4011, subregions 10 and 13 are major source areas for the acid sources. Subregion 11 may be a source region. All other loadings on endpoint variables are negative or near zero. For station 1011, subregions 12, 13, 10, 14 have relatively large loadings, but they are weaker than at the other stations. These results suggest that the Ohio River Valley region is the major source area of the acid sources in the precipitation of southern Ontario. The effect of this source region is stronger on Dorset, moderate on Kingston and weaker on London, the site closest to the source region. This effect sequence may be understood by considering the prevailing wind direction. Because the circulation pattern is from southwest to northeast, the wind will not carry as much to the west (London) as to the northeast of the eastern Ohio River valley source area (Dorset and Kingston).

Factor 2 is Ca and Mg sources. According to the results of station 1011, subregions 9 and 12 seem to be the source areas for these sources. These two subregions also show significant loadings at the other two stations. Other loadings with values of approximately 0.1 are found for subregions 7 and 10 at two stations each. Subregions 2 and 3 have such values at only one station each. These sources thus seem to be more widely spread than the acid sources, and their sources are more likely to be in Illinois, Indiana and part of Kentucky and Missouri. This region is an area of intense row crop agriculture and thus has the strong potential for wind erosion of soil. The soil particles can be incorporated into cloud droplets and the Ca and Mg leached from them in transit.

The results suggest that the Na and Cl sources are from the Gulf of Mexico, Atlantic Ocean, and some from the Pacific Ocean. The largest loadings are on subregions 14 and 4 for station 3011, on subregions 3, 13, 1, 9, and 6 for station 4011, and on subregions 13, 11 and 4 for station 1011. These subregions fall into two groups: one is subregions 3, 11, 14, and 13 where the Na and Cl sources, Atlantic or the Gulf marine aerosol can affect the system; another one is subregions 4, 5, 6, 1, and 9. In many cases during the year, particularly in winter, the air crosses the North American continent from the west coast to the east coast. In this way, the Pacific maritime aerosol can be incorporated into the cloud droplets and carried to the eastern part of the country. Since the data are only two-day back trajectories, the trajectories do not go further than about 110° longitude. This limitation makes it appear that the source area is around Iowa, Missouri, Nebraska, and Kansas. Comparing the loadings, it seems that the Gulf marine aerosol produces stronger effects.

POTENTIAL SOURCE CONTRIBUTION FUNCTION (PSCF)

Another approach to investigate the location of the sources of the observed constituents is the Potential Source Contribution Function (Malm *et al.*, 1986). Again the possible source region is subdivided into a grided i by j array. Let N represent the total number of trajectory segment endpoints during the whole study period, T. If n_{ij} endpoints fall into the ij-th cell, the probability of this event, A_{ij}, is given by

$$P[A_{ij}] = \frac{n_{ij}}{N} \qquad (2)$$

$P[A_{ij}]$ represents the residence time of a randomly selected air parcel on the ij-th cell relative to time period T. In the same ij-th cell, if there are m_{ij} endpoints that correspond to the trajectories that arrived at a receptor site with pollutant concentrations higher (or lower) than some pre-specified value, then the probability of this event, B_{ij}, is

$$P[B_{ij}] = \frac{m_{ij}}{N} \tag{3}$$

$P[B_{ij}]$ refers to the residence time for these contaminated air parcels. The potential source contribution function (PSCF) is then defined as a conditional probability:

$$PSCF = \frac{P[B_{ij}]}{P[A_{ij}]} = \frac{m_{ij}}{n_{ij}} \tag{4}$$

PSCF is the probability that an air mass with specified pollutant concentrations arrive at a receptor site after having been observed to reside in a specific geographical cell. Cells containing pollutant sources will have high conditional probabilities. Therefore, the conditional probability function, PSCF, will identify those source areas that have a potential to contribute to the high concentrations of contaminants observed at the receptor site (Malm *et al.*, 1986).

PSCF Based on Factor Scores

For each precipitation monitoring station, the PSCF values can be calculated for each 1° longitude by 1° latitude cell for each factor by specifying that the factor score for a given sample event must be greater than 0. This specification means that the contribution of the particular source related to the specified factor for the sample is higher than average since the factor score is standardized to an average value of 0. For any cases in which the factor score is greater than 0 (higher than average), the corresponding trajectory endpoints are counted in the m_{ij} in equation (3). In this way, PSCFs for each factor at each station are calculated and then plotted in Figures 6-8. The PSCF plots show a similar pattern among three stations so that only the plots for station 3011 are presented here.

In the calculation of the PSCF values, some grid cells will only have 1 endpoint (n_{ij} in equation (4) is 1). If this endpoint happens to correspond to a pollution event trajectory, the PSCFs for these cells will be 1, but the confidence in these PSCFs are very low. In this work, a

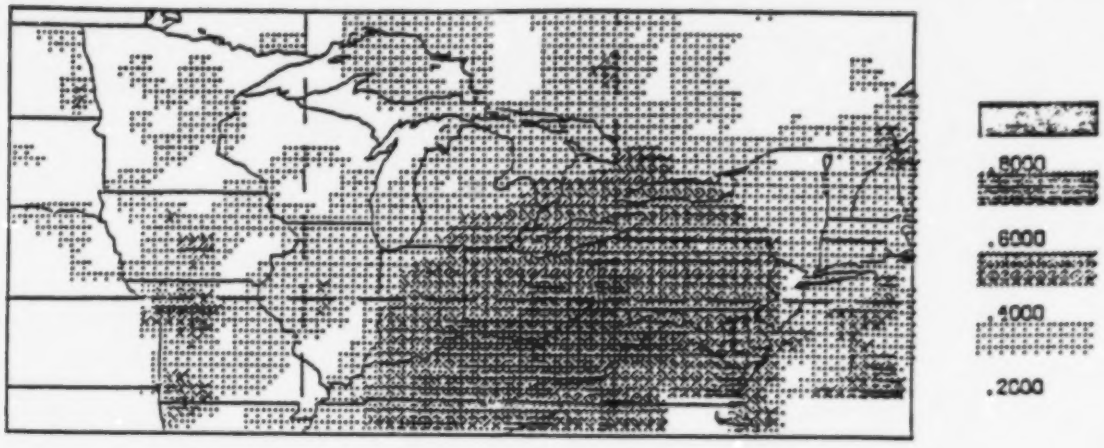

Figure 6. PSCF for factor 2 at station 3011.

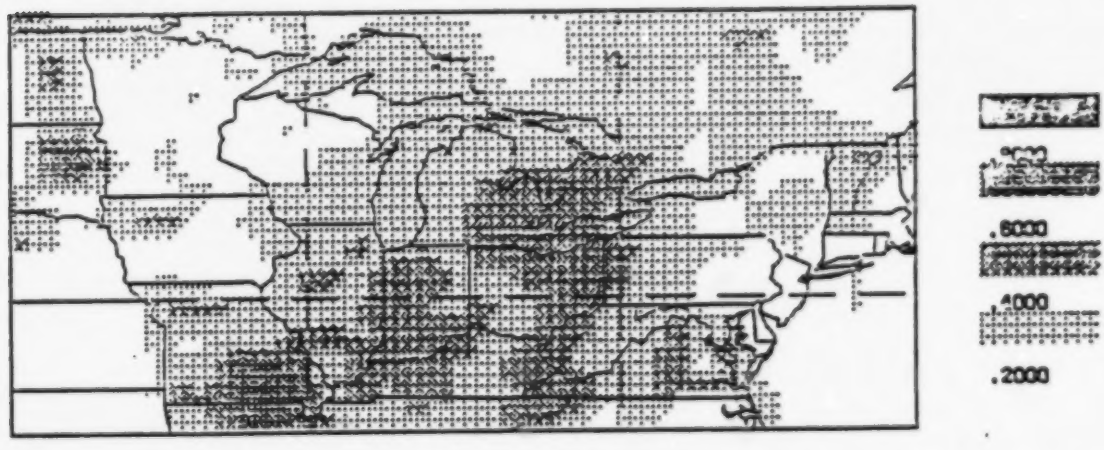

Figure 7. PSCF for factor 2 at station 3011.

weight of 0.5 is given to PSCF for the case of n_{ij} = 1, a weight of 0.68 for n_{ij} = 2, a weight of 0.85 for n_{ij} = 3, and a weight of 1.0 for $n_{ij} \geq 4$. These weighing values are of arbitrary settings.

The potential contribution probabilities for the acid sources (factor 1) are above 0.6 in a wide area of the midwest and along the east coast. Particularly, the PSCFs are higher than 0.8 along the Ohio River Valley and some areas in the east coast. There are some high density spots in the western part of Missouri. It seems that these spots are shifted westward from the St. Louis area (a high NO_x, NH_3, and SO_x emission area) possibly due to the increasing errors in the trajectories as the distance from the receptor site increases. The distribution of the dark areas in the PSCF plots for Ca and Mg sources (factor 2) looks more widely spread around the midwest region, but still concentrated a little more on the Ohio River and the Mississippi River Valleys. The distribution for Na and Cl sources (factor 3) are even more dispersed, but seldom in the area north of the stations. These sources are considered to be marine aerosol that could be carried from the Gulf, the Atlantic, or the Pacific areas as discussed in the previous section.

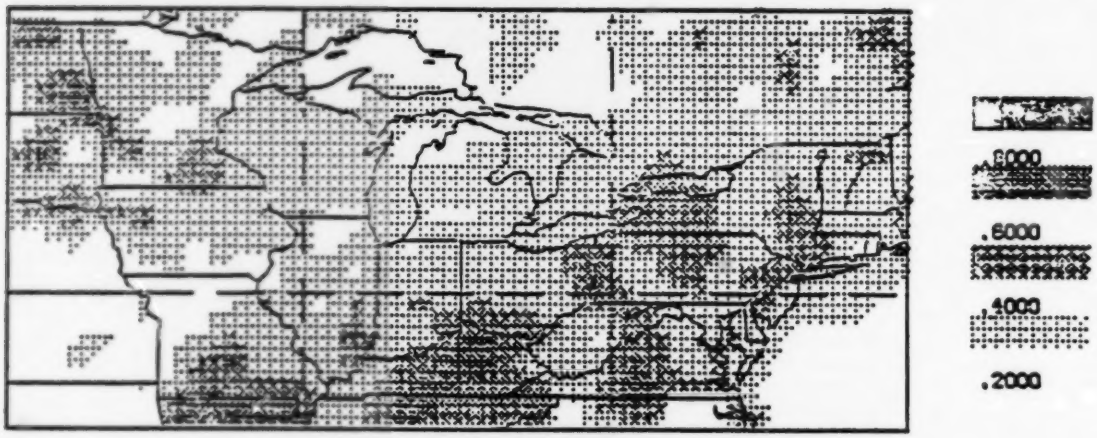

Figure 8. PSCF for factor 3 at station 3011.

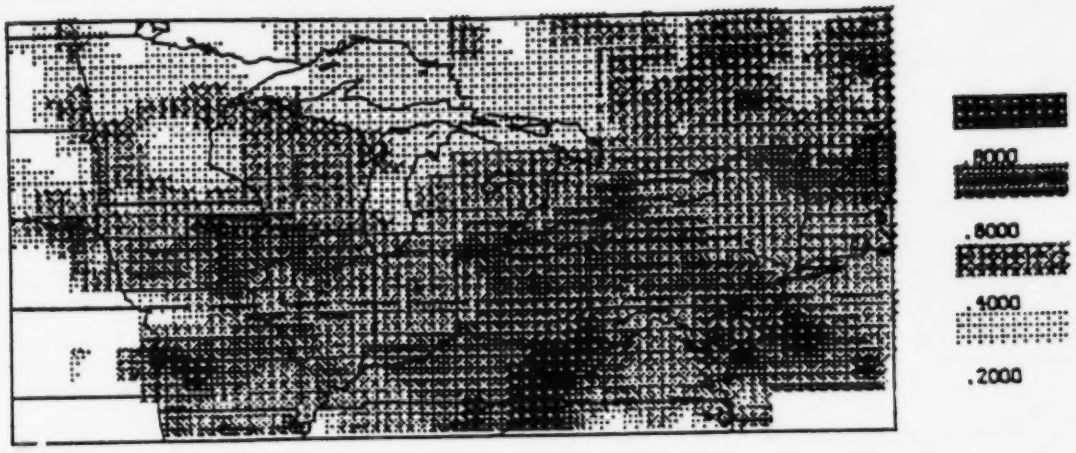

Figure 9. PSCF based on H^+ deposition at station 3011.

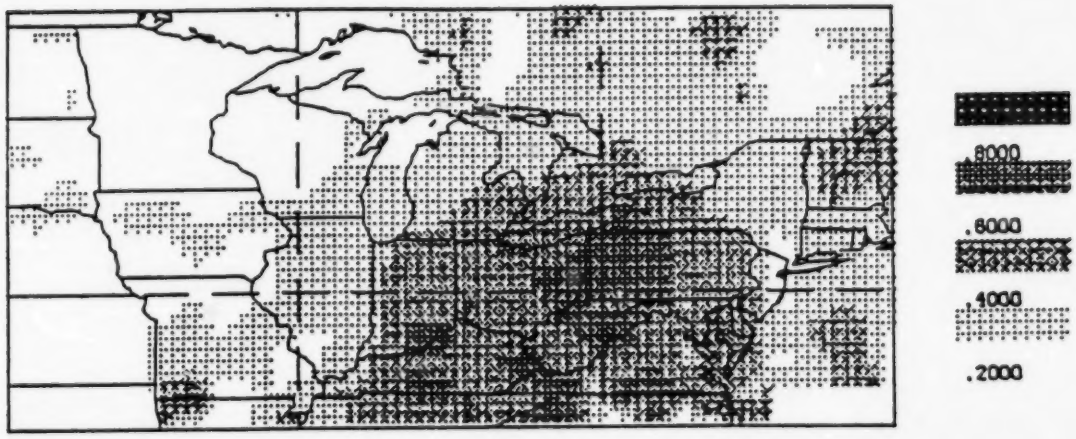

Figure 10. PSCF based on $SO_4^=$ at station 3011.

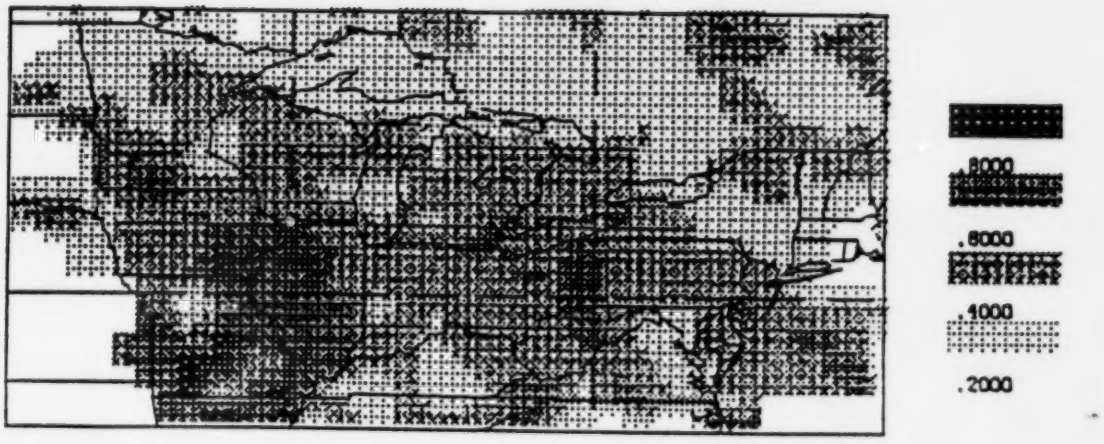

Figure 11. PSCF based on NO_3^- at station 3011.

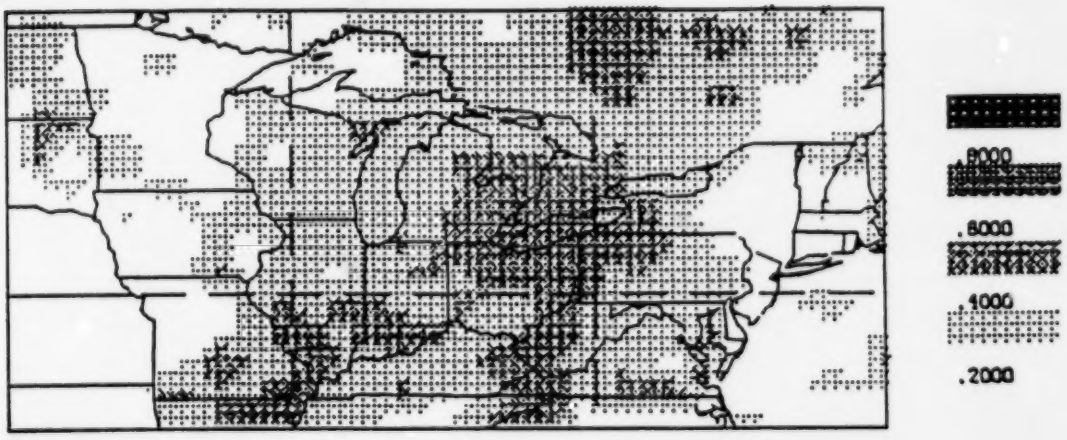

Figure 12. PSCF based on Ca^{2+} at station 3011.

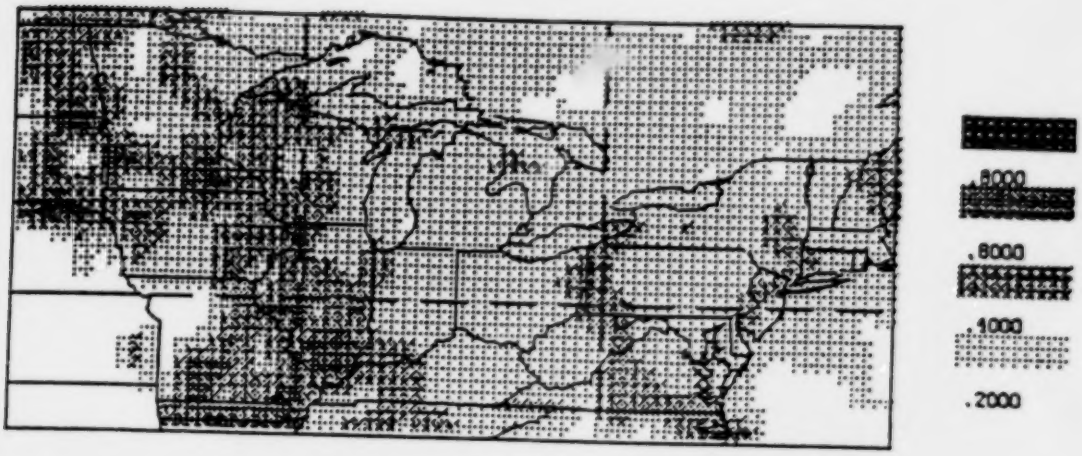

Figure 13. PSCF based on Na^+ at station 3011.

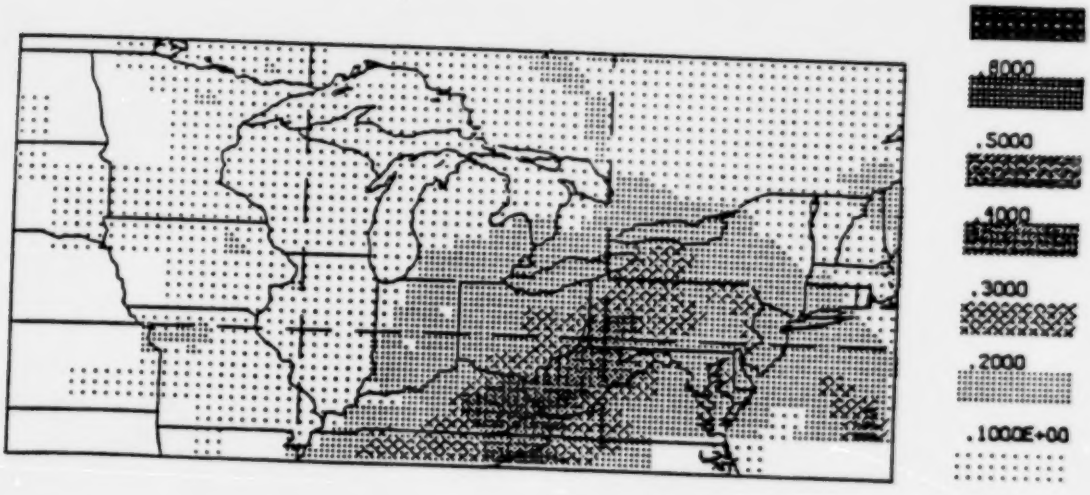

Figure 14. Joint probability plot for stations 3011 and 4011 based on factor 1 scores.

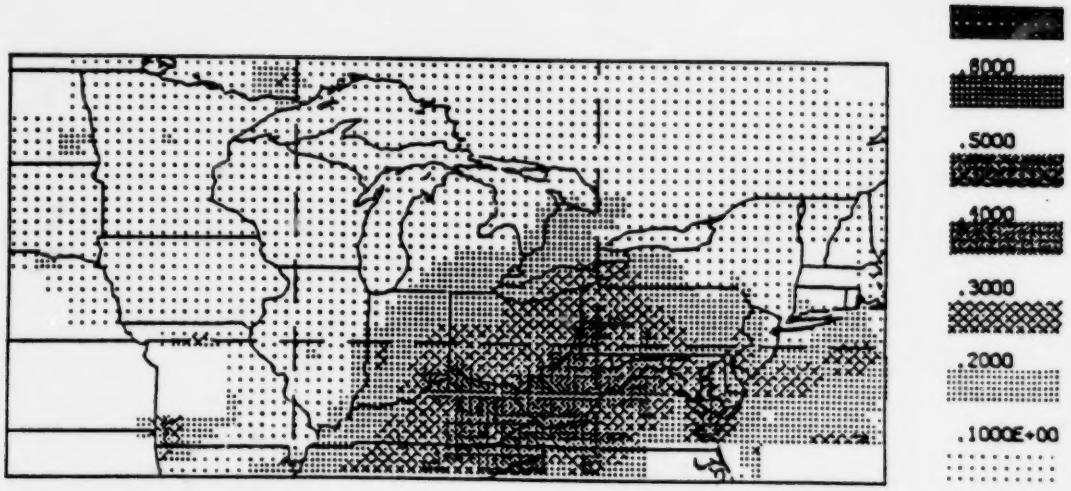

Figure 15. Joint probability plot for stations 1011 and 4011 based on factor 1 scores.

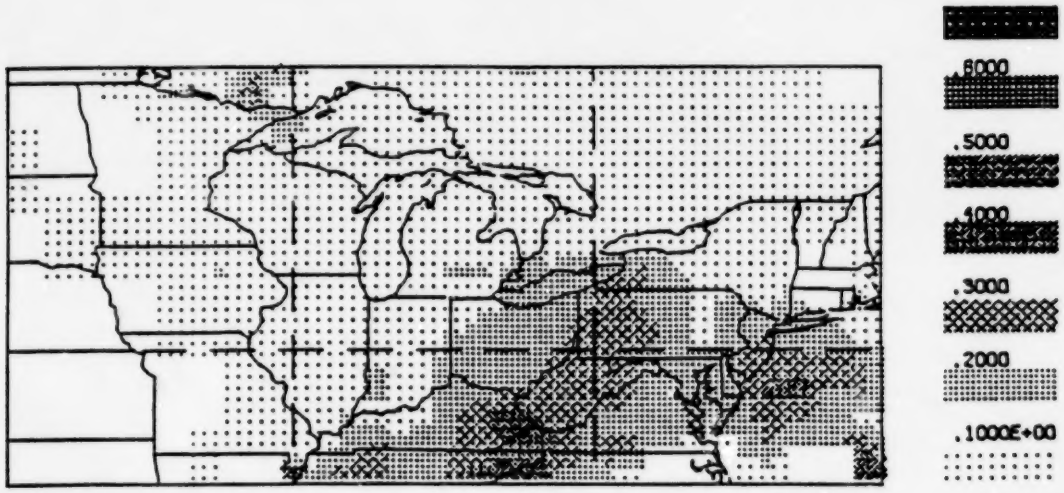

Figure 16. Joint probability plot for stations 3011 and 1011 based on factor 1 scores.

PSCF Based on Individual Species

PSCF also can be calculated on an individual species basis instead of factor scores. For example, the PSCF for pH (or H) can be calculated in each cell by specifying pH less than the average pH value (instead of factor score greater than 0). These kinds of PSCF plots may provide some more specific information. Figures 9-13 are the PSCF plots for station 3011 for pH, SO_4^{2-}, NO_3^-, Ca^{2+}, and Na^+ with the specified values (average) of 4.36, 2.45 mg/l, 0.565 mg N/l, 0.208 mg/l, and 0.051 mg/l, respectively. The PSCF of pH is similar to the PSCF for factor 1. It reflects that the potential contribution areas of the precipitation acidity are basically the Ohio River Valley, the east coast and the St. Louis region. The major contributors to the acidity (pH) are SO_4^{2-} and NO_3^-. From Figures 10 and 11, it can be seen these two acid species come from different areas. SO_4^{2-} comes from the Ohio River Valley. However more of NO_3^- comes from St. Louis and Chicago areas. The area around Pittsburgh also contributes some of the NO_3^-. These results are consistent with the distribution of emissions of SO_2 and NO_x. Figure 12 shows that Ca^+ comes from the midwest, but the probabilities are not so high. Figure 13 suggests more Na^+ comes from the Pacific. The PSCFs of Mg^+ and Cl^- are similar to that of Ca^+ and Na^+, respectively. Similar results have been obtained for the other two stations.

Joint Probability

The joint probability of potential source contributions to two receptors can be calculated and plotted out simply by multiplying the PSCFs of two stations cell by cell. It may show the common source contribution areas to both stations. Figures 14-16 are the joint probability plots for acid sources between the stations. These plots emphasize that the Ohio River Valley is the major source area of acid precipitation in southern Ontario.

CONCLUSIONS

The data analysis on the acid precipitation data sets from Ontario shows that there are three principle components contributing to the precipitation chemistry. Based on their chemical nature, these three components can be referred as acid gas sources, Ca and Mg sources, and Na and Cl sources. They are related to SO_x and NO_x emission, soil particles and marine aerosol, respectively. The acid sources are important and responsible for the acidity of the precipitation. By incorporating the information of air parcel trajectory, the acid sources can

be located in the Ohio River Valley and the east coast, the Ca and Mg sources in the midwest, particularly in the converging area of the Ohio and the Mississippi river, and Na and Cl sources from the Gulf, Atlantic, and Pacific areas. These results suggest that the Sudbury area is not a source area. The reasons might include the prevailing wind direction (southwest), high stack and relatively short distance from the stack to the monitoring stations.

The approach of incorporating trajectory information into factor analysis in the receptor modeling was originally used in particulate matter study (Malm *et al.*, 1986). It seems more difficult to use this approach in precipitation studies since precipitation involves additional chemical and physical processes. Particles have a tendency to better maintain their characteristics when compared with precipitation. More factors affect the precipitation composition during the formation and transport within the distance of the trajectory. For instance, if another precipitation event occurs in transit, then much of the material can be washed out and the pollutant concentration, e.g. $SO_4^=$ will be reduced even though the trajectory has passed an SO_2 emission area. In all cases, there are several trajectories that represent air movement for the same sampling period. In these trajectories, some of them may pass through clean areas, but others pass through source areas. They all correspond to the same precipitation sample. This mixing of air parcels will interfere with separating the regional influences and provide uncertain results. For this reason, a shorter sampling period may help reduce the interference and increase the precision of these analyses. In addition, if multilayer trajectories are used, the results may be improved.

Local Scale Particle Source Apportionment

INTRODUCTION

Ambient air quality standards for total suspended particles (TSP) created the need to identify particle sources so that effective control strategies could be designed and implemented. The initial efforts that are generally made at the identification of particle sources focused on dispersion models of point sources and in most cases, resulted in substantial reductions in TSP levels. However, as the increment of additional control needed to reach standard levels became smaller, the model uncertainties lead to difficulties in identifying the actual sources of continuing problems. In addition, fugitive and other non-ducted emissions are generally not

treated or are poorly handled in these models. Thus, additional methods were required to identify and quantitatively apportion particle mass to sources.

Again, the measured properties of the collected ambient samples are used to infer the contributions of the sources to the ambient pollutant concentration. These methods therefore require that samples be obtained at locations of interest, receptor sites, and that the samples so collected be analyzed for the properties that are characteristic of the pollutant sources.

These requirements have arisen at a time when new analytical methods have been developed that permit multielemental analysis of large numbers of airborne particle samples. Thus, large data bases on the composition of airborne particles are available for use in these receptor models. Although much of the thrust of the model developments have been aimed at identification of sources of particle mass, they also can be used to elucidate the origins of the various measured species observed in the samples. It then becomes possible to quantitatively apportion the observed airborne concentrations such as airborne lead among the various source types.

The importance of receptor models as air quality management tools in the United States has recently substantially increased because of the promulgation of a new ambient air quality standard for particulate matter. This new standard requires all of the state and local air quality planning agencies to revise their plans for improving air quality and reducing the particulate level concentrations where they are expected to exceed the prescribed levels. In the associated guidance documents provided by the U.S. Environmental Protection Agency (1986), receptor models are explicitly approved for use in this planning process along with the traditional dispersion models. Thus, receptor models have now become an accepted tool for air quality management. Several of the applicable models and examples of their use in apportioning particle mass to sources will be presented.

PRINCIPLE OF MASS CONSERVATION

All of the currently used receptor models are based on the concept of conservation of mass and the use of a mass balance analysis. For example, let us assume that the total airborne particulate lead concentration (ng/m^3) measured at a site can be considered to be the sum of contributions from independent source types such as motor vehicles, incinerators, smelters, etc.

$$Pb_T = Pb_{auto} + Pb_{incin.} + Pb_{smelter} + \ldots \quad (5)$$

However, a motor vehicle burning leaded gasoline emits particles containing materials other than lead. Therefore, the atmospheric concentration of lead from automobiles in ng/m^3, Pb_{auto}, can be considered to be the product of two cofactors; the gravimetric concentration (ng/mg) of lead in automotive particulate emissions, $a_{Pb,auto}$, and the mass concentration (mg/m^3) of automotive particles in the atmosphere, f_{auto}.

$$Pb_{auto} = a_{Pb,auto} f_{auto} \quad (6)$$

The normal approach to obtaining a data set for receptor modeling is to determine a large number of chemical constituents such as elemental concentrations in a number of samples. The mass balance equation can thus be extended to account for all m elements in the n samples as contributions from p independent sources

$$x_{ij} = \sum_{k=1}^{p} a_{ik} f_{kj} \qquad \begin{matrix} i=1,m \\ j=1,n \end{matrix} \quad (7)$$

where x_{ij} is the ith elemental concentration measured in the jth sample, a_{ik} is the gravimetric concentration of the ith element in material from the kth source, and f_{kj} is the airborne mass concentration of material from the kth source contributing to the jth sample. There are several different approaches to receptor model analysis that have been successfully applied including chemical mass balance (CMB), and multivariate receptor models including principal components analysis and target transformation factor analysis (TTFA). These models can be applied to both particulate and gaseous species. The basis for each of these methods will be presented in subsequent sections with examples of their application to the identification of particle sources in Hamilton.

CHEMICAL MASS BALANCE

The chemical mass balance (CMB) sometimes called the chemical element balance solves equation 7 directly for each sample by assuming that the number of sources and their compositions at the receptor site are known. This approach was first independently suggested by Winchester and Nifong (1971) and by Miller, Friedlander, and Hidy (1972). The composition

of an ambient sample is then used in a multiple linear regression against source compositions to derive the mass contribution of each source to that particular sample.

It must be made clear, however, that the CMB analysis works well in case when both the source and ambient samples were collected and analyzed during the same time period. Much less detailed resolution of sources are generally possible when its sources are not sampled and analyzed. In an intercomparison study organized by the U.S. Environmental Protection Agency (Stevens and Pace, 1984) to examine receptor models, a set of ambient particulate elemental compositional data sets were analyzed by a number of investigators using similar CMB methods. The compositions of particles from sources in Houston were not available and were not measured during this program so that source composition profiles had to be obtained from literature sources. The lack of source data immediately raised problems in the use of the mass balance methods and comparison of results from different investigators (Dzubay *et al.*, 1984). It is not always certain exactly which sources should be included in the analysis. Although emission inventories may be available for the region, it may be that the measured source composition for a coal-fired power plant in Maryland burning eastern bituminous coal is not a particularly good representation for a plant in Hamilton.

An additional problem for receptor modeling is that profiles of real sources change their characteristics in time. For example, the motor vehicle profile is undergoing rapid changes in lead and bromine concentrations with time as the mix of new, catalyst-equipped and diesel cars and leaded-fuel burning vehicles change.

MULTIVARIATE RECEPTOR MODELS

Alternative approaches have been developed for identifying and quantitatively apportioning sources of airborne particles using multivariate statistical analysis. Eigenvector analysis has been the principal method that has been applied to airborne particle composition data. An eigenvector analysis tries to simplify the description of a system by determining the minimum number of new variables necessary to reproduce the measured attributes of the system. The mathematical basis of these methods has been described by Hopke (1985).

Principal components and factor analysis are names given to several of the variety of forms of eigenvector analysis. It was originally developed and used in psychology to provide mathematical models of psychological theories of human ability and behavior (Harman, 1976).

However, eigenvector analysis has found wide application throughout the physical and life sciences. Unfortunately, a great deal of confusion exists in the literature in regard to the terminology of eigenvector analysis. Various changes in the way the method is applied has resulted in it being called factor analysis, principal components analysis, principal components factor analysis, empirical orthogonal function analysis, Karhunen-Loeve transform, etc., depending on the way the data are scaled before analysis or how the resulting vectors are treated after the eigenvector analysis is completed. All of the methods have the same basic objective; the compression of data into fewer dimensions and the identification of the structure of interrelationships that exist between the variables measured or the cases studied.

The first step in the eigenvector analysis is the calculation of a dispersion matrix, the matrix that contains quantitative information on the relative variation of pairs of variables or pairs of samples (cases). There are two basic types of dispersion matrices. They are covariance matrices and correlation matrices. For a correlation matrix, the data are scaled such that each variable or each case has an equal weight. The data are not scaled before calculating covariance. In both instances, the data may be centered by subtracting a mean value before scaling and the calculation of the matrix elements. The choice of dispersion matrix depends on the nature of the data set to be analyzed. For many types of chemical spectroscopic data, the covariance matrix is the choice because each variable has the same measurement scale. For many geochemical problems, the difference in scale between major, minor, and trace components requires scaling to avoid domination of the analysis by the major components.

The dispersion matrix is then decomposed into a series of orthogonal vectors by the process outlined by Joreskog, Klovan, and Reyment (1976) so that

$$U'DU = \Lambda \tag{8}$$

where U is the matrix of eigenvectors, U' is its transpose, D is the dispersion matrix, and Λ is a diagonal matrix of eigenvalues where the trace of Λ is equal to the trace of D. If there were no errors in the data from which D is calculated, the number of non-zero eigenvalues would be the dimensionality of the problem called the rank of D. The rank for the original data matrix is the same as that for the dispersion matrix. However, experimental error generally results in a number of small but non-zero eigenvalues. The determination of the number of vectors containing significant information relative to those dominated by noise is often a difficult one.

The lack of universally applicable criteria for determining the dimensionality of the data is a major problem in the application of factor analysis.

In the most commonly used approach to calculating the eigenvectors, the maximum amount of variance is packed into the first eigenvalue. The maximum possible amount of the remaining variance goes into the second and so forth. This compression of the information into a few components permits much of the variation in the data set to be displayed in a two or three dimensional plot. For many classification problems, the first few factors are able to reproduce most of the data structure and to remove some of the noise. The objects can then be plotted using the components axes and thus display the features of high dimensional data in a few dimensions (Blackith and Reyment, 1971).

The compression of variance into the first factors will improve the ease with which the number of factors can be determined. However, their nature has now been mixed by the calculational method. Thus, once the number of factors has been determined, it is often useful to rotate the axes in order to provide a more interpretable structure. The axis rotation can retain the orthogonality of the eigenvectors or cause them to be oblique. Depending on the initial data treatment, the axes rotations may be in the scaled and/or centered space or in the original variable scale space. The latter approach has proved quite useful in a number of chemical applications described by Malinowski and Howery (1980) and in environmental systems as described by Hopke (1985).

One of the valuable uses of this type of analysis is in screening large data sets to identify errors (Roscoe *et al.*, 1982). With the use of atomic and nuclear methods to analyze environmental samples for a multitude of elements, very large data sets havee basic objective: the compression of data into fewer dimensions and the identification of the structure of interrelationships that exist between the variables measured or the cases studied. Principal component factor analysis can provide useful insight into several possible problems that may exist in a data set including incorrect single values and some types of systematic errors.

In the early applications of factor analysis to particulate compositional data, it was generally easy to identify a fine particle mode lead/bromine factor that could be assigned as motor vehicle emissions. In many cases, a calcium factor sometimes associated with lead could be found in the coarse mode analysis and could be assigned as road dust. However, the problem of diminishing lead concentrations in gasoline represents a problem to identify the

influence of motor vehicles. As the lead and related bromine concentrations diminish, the clearly distinguishable covariance of these two elements disappears. In a study of particle sources in southeast Chicago based on samples from 1985 and 1986, much lower lead levels are observed and the lead/bromine correlation is quite weak (Hopke *et al.*, 1988). Thus, the identification of highway emissions through factor analysis based on lead or lead and bromine is becoming more and more difficult and other species are going to be needed in the future.

A problem that exists with these forms of factor analysis is that they do not permit quantitative source apportionment of particle mass or specific elemental concentrations. In an effort to find an alternative method that would provide information on source contributions when only the ambient particulate analytical results are available, Hopke and coworkers have developed target transformation factor analysis (TTFA). This approach was recently reviewed by Hopke (1988). In this analysis, resolution similar to that obtained from a CMB analysis can be obtained. However, a CMB analysis can be made on a single sample if the source data is known while TTFA requires a series of samples with varying impacts by the same sources, but does not require *a priori* knowledge of the source characteristics.

In matrix notation equation 3 can be rewritten as

$$X = AF$$

where X is the matrix of ambient aerosol compositions, A is the matrix of source composition profiles, and F is the matrix of mass contributions of the sources to the samples. The objectives of TTFA are 1) to determine p, the number of independent sources that contribute to the system, 2) to identify the components of matrix A, the elemental source profiles, and 3) to calculate F, the contribution of each source to each sample.

One of the first applications of TTFA was to the source identification of urban street dust (Lamb *et al.*, 1980). The sample of street dust was vacuumed from an intersection in Urbana, Illinois. The sample was physically fractionated by particle size, density, and magnetic susceptibility to produce 30 subsamples. Each subsample was analyzed by instrumental neutron activation analysis and atomic absorption spectroscopy to yield analytical results for 35 elements.

The number of sources is determined by performing an eigenvalue analysis on the matrix of correlations between the samples. A target transformation determines the degree of overlap between an input source profile and one of the calculated factor axes. The input source profiles, called test vectors, are developed from existing knowledge of the emission profiles of

various sources or by an iterative technique from simple test vectors (Roscoe and Hopke, 1981). The identified source profiles are then used in a simple weighted least-squares determination of the mass contributions of the sources (Severin *et al.*, 1983).

In the analysis of the street dust, six sources were identified including soil, cement, tire wear, direct automobile exhaust, salt and iron particles. The lead concentration of the motor vehicle source was found to be 15% with a lead to bromine ratio of 0.39. This ratio is in good agreement with the values obtained by Dzubay *et al.* (1979) for Los Angeles freeways and in the range presented by Harrison and Sturges (1983) in their extensive review of the literature. One of the principal advantages of TTFA is that it can identify the source composition profiles as they exist at the receptor site. There can be changes in the composition of the particles in transit from the source to the receptor and approaches that provide these modified source profiles should improve the receptor model results. Chang *et al.* (1988) have applied TTFA to an extensive set of data from St. Louis, MO, to develop source composition profiles based on a subset selection process developed by Rheingrover and Gordon (1988).

Rheingrover and Gordon (1988) identify samples strongly affected by single point sources using wind-trajectory analysis. They select samples from a data base such as the one obtained in the Regional Air Pollution Study (RAPS) of St. Louis, MO, that were heavily influenced by major sources of each element. These samples were identified according to the following criteria:

1. Concentration of the element in question $X > X + Z_{\sigma}$ where X is the average concentration of that particular element for each station and size fraction (coarse or fine particle size fraction), Z_{σ} is typically set at about three for most elements, and is the standard deviation of the concentration of that element.

2. The standard deviation of the 6 or 12 hourly average wind directions for most samples, or minute averages for 2-hour samples, taken during intensive periods is less than 20 degrees.

Samples that are strongly affected by emissions from a source were identified through observation of clustering of mean wind directions for the sampling periods selected with angles pointing toward the source.

TARGET TRANSFORMATION FACTOR ANALYSIS

For many problems it is of interest to interpret the data in terms of a mass balance. It is assumed that the measured environmental property is a linearly additive sum of independent contributions from each of the sources. Thus, we can rewrite equation 2 such that the actual value of x_{ij} rather than the standardized value is apportioned.

$$x_{ij} = \sum_{k=1}^{p} a_{ik} f_{kj} \qquad (9)$$

where now a_{ik} is the concentration of the ith element in particles from the kth source and f_{kj} is the contribution of particle mass by the kth source to the jth sample. Principal components or empirical orthogonal vector analysis are able to identify the interrelationships between samples, locations, or time. They are able to provide correlations among the variables and causal factor, but not direct estimates of the contribution of sources to the measured aerosol mass in $\mu g/m^3$. It is the purpose of target transformation factor analysis to make such an apportionment based only on the ambient concentration data. The methodology is described by Hopke (1985; 1988; 1989) and has been successfully applied to a number of urban aerosol mass apportionment problems, artificial data to test the method, and the determination of mineral matter content in coal where the TTFA results could be compared to those obtained by x-ray diffraction. The results presented below were obtained using the TTFA code FANTASIA (Hopke *et al.*, 1983; Hopke and Dharmavaram, 1986).

SOURCE APPORTIONMENT IN HAMILTON

To examine how to apply TTFA to the source apportionment of airborne particulate matter, a data set obtained by the x-ray fluorescence (XRF) analysis of samples collected with a dichotomous sampler in Hamilton during the period of September 19, 1983 to July 7, 1986 will be examined. The dichotomous sampler provides two samples for each 24 hour sampling interval; one sample in the aerodynamic diameter range of 2.5 μm to 10 μm and the other with

aerodynamic diameter < 2.5 μm. The larger size range sample is called the "coarse" fraction and the smaller size mode is referred to as the "fine" fraction.

First, the Rheingrover and Gordon trajectory analysis approach was employed to identify particle sources and develop specific source profiles for the Hamilton airshed. We obtained the meteorological data for several sites in Hamilton from the Canadian Climate Center. The hourly average wind speed and direction values were extracted and matched to the particle sampling intervals. The meteorological data from the location nearest to the sampling site was used for this analysis. Particle composition data from all of the sites was evaluated. The data were then examined to determine if there were samples showing high elemental concentrations and constant wind direction. Unfortunately, the combination of a small set of samples and nature of the meteorology of Hamilton did not provide any data sets of more than 2 to 3 samples. The proximity to Lake Ontario results in land/sea breeze behavior and the presence of an escarpment separating the lower and upper portions of the city give rise to wind directions that do not remain constant over 24 hour periods. Thus, the trajectory selection method could not be applied to obtain sets of samples significantly affected by single point sources as had been found in the St. Louis metropolitan area. Thus, we began the TTFA examination of the entire set of samples for both the fine and coarse fractions.

The TTFA process requires data sets that have complete data for all elements in all samples. Thus, samples in which a majority of the data points were below detection limits or those variables for which a majority of the results were undetectable were eliminated. In addition the iterative TTFA procedure also requires knowledge of the total mass concentration value of each sample. For these samples, there had been difficulties in the weighing procedures (Chan, 1988). It is possible that in many cases the radioactive source used as a static eliminator to remove the charge on the filters was not employed or had decayed to the point where it would no longer produce an adequate bipolar charging field. Thus, many of the mass values of the filters are suspect. We have attempted to select as large and complete a data set as possible.

A second difficulty with these data is the absence of quantitative estimates of uncertainty in the measured elemental concentrations. The samples were analyzed by XRF. As part of that analysis, it is possible to estimate the statistical precision in the fluoresced x-ray intensities. By combining these results with the values from the blank filters and estimated uncertainties in

the volume of air that was sampled, a quantitative estimate of the uncertainty in the concentration could be obtained. It is strongly recommended to the Ministry that they develop the procedures to incorporate error analysis into the XRF analytical procedures so that the data quality can be fully assessed before further analysis is attempted. It is important to know which values are close to the limits of detection and thus have large inherent uncertainties in them. Furthermore, we have found that the squared average error is typically the most useful weight to use in the iterative target transformation vector development process. Since these errors are not available here, we have used the variances of the variables as the rotation weights. Finally, for those cases where the reported values were below detection limits, we have replaced the detection limit with a value equal to the product of the detection limit with a number between 0 and 1 selected with a uniform random number generator.

Fine Fraction Results

A total of 132 pairs of samples were obtained at site 29025, the site with the longest record of sampling. The other sites in Hamilton had too few samples to permit multivariate analysis methods to be used (Henry *et al.*, 1984). After screening the data for missing values and for designations of questionable analytical results, we have used 94 samples and 21 elemental concentrations for the fine fraction sample data set.

In order to determine the number of resolvable sources, an eigenvalue analysis is made. The results of the eigenvalue analysis are presented in Table XI. The eigenvalue provides a measure of the information content of the corresponding eigenvector. The other parameters presented in the table provide measures of the quality of the data reproduced with a given number of factors relative to the original data set. Thus, we want to choose a number of factors such that there is still a significant eigenvalue, but where the values of the data reproduction indicators have become low. In this case, an initial choice of 6 factors is made. There is a significant decrease in the eigenvalue between 6 and 7 factors and there is not a corresponding decrease in the values of the reproduction quality indicators. It is generally not possible to resolve more than about 6 to 7 sources in this type of analysis unless the noise level in the data is very low and there is considerable orthogonality in the source profiles.

The next step is the iterative development of the source profiles. Beginning from the 21 unique vectors and iterating to convergence, a set of final vectors has been obtained and are

Table XI. Results of eigenvalue analysis and data reproduction tests for fine fraction data at Hamilton site 29025.

Factor	Eigenvalue	RMS	Chi-Square*	Exner	Indicator	Average Error
1	8.0864E+01	9.4486E-03	1.8703E-01	.418343	2.0897E-04	978.12
2	7.4554E+00	4.8945E-03	5.3404E-02	.275098	1.5622E-04	763.77
3	2.2110E+00	3.6658E-03	3.1969E-02	.214989	1.3975E-04	717.77
4	1.3004E+00	2.5751E-03	1.6888E-02	.169986	1.2747E-04	669.61
5	9.0204E-01	1.7987E-03	8.8530E-03	.129912	1.1336E-04	539.58
6	7.5265E-01	1.3137E-03	5.0947E-03	.082760	8.4863E-05	484.53
7	2.2998E-01	1.1992E-03	4.6006E-03	.061523	7.4963E-05	314.43
8	9.7348E-02	9.5718E-04	3.1933E-03	.049880	7.3148E-05	310.65
9	7.2979E-02	6.5909E-04	1.6595E-03	.038932	6.9741E-05	307.96
10	5.3833E-02	3.9981E-04	6.7412E-04	.028258	6.2919E-05	302.11
11	1.7576E-02	3.3958E-04	5.4137E-04	.023755	6.7126E-05	309.60
12	1.1989E-02	2.9041E-04	4.4530E-04	.020114	7.3965E-05	284.13
13	1.0466E-02	2.5991E-04	4.0621E-04	.016283	8.0378E-05	213.25
14	6.7278E-03	2.1031E-04	3.0778E-04	.013248	9.1310E-05	153.60
15	4.5323E-03	1.8903E-04	2.9374E-04	.010729	1.0872E-04	98.74
16	3.9014E-03	1.2522E-04	1.5668E-04	.007946	1.2701E-04	73.53
17	2.1127E-03	9.0470E-05	1.0355E-04	.005915	1.6517E-04	40.15
18	1.4441E-03	6.5687E-05	7.3743E-05	.003968	2.2746E-04	23.64
19	7.2716E-04	3.9589E-05	4.0715E-05	.002461	3.8877E-04	9.13
20	2.8266E-04	2.3306E-05	2.8603E-05	.001514	1.3526E-03	3.75

*Chi-square not weighted by errors

presented in Table XII. There are a number of the iterated profiles that have quite similar characteristics. In order to group the vectors into a small number of possible source types, a cluster analysis is performed. An agglomerative hierarchical cluster analysis is performed using squared Euclidean distance and average linkage as the clustering criterion (Massart and Kaufman, 1983). The dendrogram for this cluster analysis is provided in Figure 17. The figure shows the relationships among the various vectors. The objective of this analysis is to quickly find groups of vectors that have similar characteristics and thus represent the various source types contributing to the system. It also identifies those vectors that are unique and are probably strongly related to a specific element.

To examine the groupings, a vertical line is envisioned in the dendrogram. For the one in Figure 17, a vertical line at the criterion value represent by M or N would separate the

Table XII. Iterated vectors from target transformation analysis of Hamilton fine particle composition data.

Element	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7	Vector 8
Al	1.0256E+04	4.3017E+03	4.2915E+03	5.6420E+03	4.5599E+03	4.5147E+03	4.3208E+03	2.5050E+03
As	8.3713E+02	2.0281E+03	2.1271E+03	1.8479E+03	7.5328E+02	1.9668E+03	2.0533E+03	0.0000E+00
Br	4.8261E+03	1.1174E+04	1.1889E+04	1.0470E+04	1.5120E+03	1.0776E+04	1.1352E+04	9.4244E+02
Ca	7.1154E+03	1.0173E+04	1.0887E+04	1.0270E+04	2.1987E+00	9.8523E+03	1.0354E+04	3.6918E+03
Cl	5.4761E+03	3.1068E+03	1.3830E+02	8.4655E+01	6.3183E+04	5.3290E+03	2.5011E+03	1.2308E+03
Cr	1.3843E+03	2.5512E+03	2.6581E+03	2.3859E+03	1.3866E+03	2.5040E+03	2.5808E+03	4.6921E+01
Cu	1.4996E+03	3.0865E+03	3.2448E+03	2.8781E+03	1.0260E+03	2.9973E+03	3.1266E+03	0.0000E+00
Fe	1.8594E+03	2.4759E+02	1.5455E+02	1.3905E+03	3.1348E+02	2.9807E+02	2.0192E+02	7.1136E+04
K	5.2119E+03	5.5332E+03	5.4116E+03	5.3553E+03	9.6764E+03	5.7332E+03	5.5254E+03	7.7493E+03
Mn	2.0553E+03	2.6948E+03	2.7929E+03	2.6841E+03	1.7284E+03	2.6752E+03	2.7227E+03	3.6766E+03
Ni	1.2985E+03	2.5998E+03	2.6869E+03	2.3741E+03	1.8258E+03	2.5615E+03	2.6250E+03	0.0000E+00
Pb	1.5743E+04	3.7813E+04	4.0279E+04	3.5405E+04	3.8168E+03	3.6388E+04	3.8415E+04	7.9213E+03
P	1.2810E+03	3.1642E+03	3.2455E+03	2.7999E+03	2.2450E+03	3.0927E+03	3.1846E+03	4.8677E+02
Rb	3.0898E+02	3.8008E+02	3.1039E+02	2.6604E+02	1.6754E+03	4.1963E+02	3.6532E+02	0.0000E+00
Se	1.2545E+03	2.4921E+03	2.6253E+03	2.3414E+03	8.3320E+02	2.4264E+03	2.5275E+03	0.0000E+00
Si	3.2496E+04	1.4438E+02	5.2574E+01	7.4790E+03	8.3492E+02	1.2140E+03	2.0362E+02	1.1600E+02
Sr	5.7085E+02	1.0637E+03	1.0681E+03	9.3772E+02	1.3315E+03	1.0681E+03	1.0671E+03	0.0000E+00
S	3.4568E+03	1.6377E+03	1.6822E+02	1.0858E+02	1.4211E+02	5.7862E+02	1.0692E+03	4.1005E+02
Ti	1.5969E+03	2.8470E+03	2.9856E+03	2.6978E+03	1.1514E+03	2.7818E+03	2.8841E+03	8.3020E+01
V	1.3466E+03	2.6513E+03	2.7671E+03	2.4604E+03	1.3578E+03	2.5964E+03	2.6829E+03	0.0000E+00
Zn	1.2592E+02	3.1018E+02	2.1698E+02	1.2251E+02	6.4445E+02	2.2647E+02	2.3664E+02	4.1781E+00

Element	Vector 9	Vector 10	Vector 11	Vector 12	Vector 13	Vector 14	Vector 15	Vector 16
Al	5.4195E+03	5.1228E+03	4.3582E+03	4.2255E+03	4.2884E+03	4.4544E+03	4.3969E+03	1.5590E+04
As	1.2106E+03	1.4223E+03	1.9407E+03	2.1140E+03	1.8944E+03	1.1732E+03	2.0634E+03	7.0105E+01
Br	6.6157E+03	8.2918E+03	1.0523E+04	1.1857E+04	1.0173E+04	4.8047E+03	1.1417E+04	1.1997E+03
Ca	6.7588E+03	8.4812E+03	9.5010E+03	1.0857E+04	9.1481E+03	3.6496E+03	1.0435E+04	6.3243E+03
Cl	1.5043E+04	5.0284E+03	7.4574E+03	6.4874E+01	7.4006E+03	3.5066E+04	2.5893E+03	6.5503E+02
Cr	1.7482E+03	1.9432E+03	2.4774E+03	2.6412E+03	2.4050E+03	1.6834E+03	2.5990E+03	6.6504E+02
Cu	1.9273E+03	2.2717E+03	2.9475E+03	3.2358E+03	2.8678E+03	1.7031E+03	3.1419E+03	5.6472E+02
Fe	1.0789E+04	1.2433E+04	3.1080E+02	9.8829E+01	2.8347E+02	4.9982E+02	1.5540E+02	7.5844E+00
K	6.6622E+03	6.0869E+03	5.8682E+03	5.3470E+03	5.7085E+03	7.3043E+03	5.5828E+03	4.4542E+03
M	2.5493E+03	2.7793E+03	2.6435E+03	2.7835E+03	2.5507E+03	1.9048E+03	2.7434E+03	1.6713E+03
Ni	1.7769E+03	1.9009E+03	2.5526E+03	2.6692E+03	2.4841E+03	1.9453E+03	2.6424E+03	4.2705E+02
Pb	2.2671E+04	2.8687E+04	3.5512E+04	4.0251E+04	3.4346E+04	1.5600E+04	3.8595E+04	3.0110E+03
P	2.2211E+03	2.3917E+03	3.0976E+03	3.2230E+03	3.0635E+03	2.5092E+03	3.1906E+03	0.0000E+00
Rb	4.5219E+02	2.1156E+02	4.6935E+02	3.0107E+02	4.8569E+02	1.1112E+03	3.6549E+02	1.0194E+02
Se	1.5697E+03	1.8478E+03	2.3829E+03	2.6058E+03	2.3088E+03	1.3538E+03	2.5462E+03	5.3019E+02
Si	7.3344E+03	6.0327E+03	3.8147E+02	2.1250E+01	5.8267E+01	7.3671E+02	4.9576E+02	6.1726E+04
Sr	7.8940E+02	7.3153E+02	1.0854E+03	1.0634E+03	1.0617E+03	1.1193E+03	1.0734E+03	1.9839E+02
S	6.3252E+02	1.0684E+02	8.8525E+02	1.1524E+02	3.9754E+03	9.4797E+03	3.1105E+02	3.6754E+02
Ti	1.8896E+03	2.1711E+03	2.7363E+03	2.9669E+03	2.6557E+03	1.6699E+03	2.9044E+03	8.8538E+02
V	1.7417E+03	1.9549E+03	2.5679E+03	2.7506E+03	2.4936E+03	1.7035E+03	2.7009E+03	5.3967E+02
Zn	1.9778E+02	1.0386E+02	3.0130E+02	8.0764E+02	3.4642E+02	5.2747E+02	5.1308E+01	1.0112E+03

Table XII (continued). Iterated vectors from target transformation analysis of Hamilton fine particle composition data.

Element	Vector 17	Vector 18	Vector 19	Vector 20	Vector 21
Al	4.3419E+03	4.1270E+03	4.6945E+03	4.3521E+03	9.6363E+02
As	1.7935E+03	4.9056E+02	1.9751E+03	2.0224E+03	0.0000E+00
Br	9.4135E+03	1.5930E+02	1.0914E+04	1.1060E+04	2.7051E+03
Ca	8.3293E+03	4.6028E-03	1.0106E+04	1.0019E+04	3.5623E+03
Cl	1.3819E+04	4.0765E-01	3.3936E+03	5.0361E+03	6.9565E+03
Cr	2.3345E+03	2.9909E+02	2.5097E+03	2.5598E+03	7.6661E+01
Cu	2.7095E+03	5.3421E+02	3.0205E+03	3.0694E+03	8.8718E+02
Fe	3.2069E+02	5.8141E+01	2.7685E+02	2.1936E+01	3.4453E+03
K	6.2599E+03	9.8198E+02	5.5838E+03	5.7397E+03	5.6000E+02
Mn	2.5174E+03	2.8372E+00	2.6872E+03	2.6980E+03	1.4405E+03
Ni	2.4524E+03	4.0693E+02	2.5503E+03	2.6219E+03	1.5679E+01
Pb	3.1638E+04	7.6640E+02	3.6873E+04	3.7326E+04	1.6134E+04
P	2.9916E+03	1.7784E+03	3.0679E+03	3.1738E+03	0.0000E+00
Rb	6.0980E+02	7.8071E+02	3.7518E+02	4.2419E+02	0.0000E+00
Se	2.1868E+03	1.9782E+02	2.4465E+03	2.4876E+03	0.0000E+00
Si	2.8419E+02	1.5452E+02	2.2007E+03	1.7934E+02	1.0868E+04
Sr	1.1078E+03	2.7032E+02	1.0455E+03	1.0896E+03	2.9780E+02
S	1.5047E+03	8.7950E+04	6.7202E+02	6.0596E+02	1.4284E+03
Ti	2.5347E+03	3.6800E+02	2.8030E+03	2.8438E+03	7.1277E+01
V	2.4112E+03	3.1193E+02	2.6024E+03	2.6582E+03	9.8694E+01
Zn	4.3921E+02	3.6137E+02	2.0185E+02	1.0502E+01	5.0489E+04

vectors into 6 groups. Vectors 7, 15, 2, 6, 19, 20, 3, 12, 11, 13, 4, 17, 9, and 10 form one large group. Vectors 5 and 14 and Vectors 1 and 16 form two smaller groups. Vectors 18, 21, and 8 are each unique vectors.

From these patterns, vectors are chosen 6 at a time to reproduce the original data set. The objective is to find a set of vectors that adequately reproduce the original data set. The solution should not produce significant numbers of negative **f** values since the mass contribution of the sources can only be positive. The program provides the opportunity to test a large number of combinations of source profiles to determine the optimum result. In this case, the combination of Vectors 4, 5, 8, 16, 18, and 21 very quickly is found to provide the best results. Using this combination of vectors, the elements of the F matrix are calculated using the least-squares method of Severin *et al.* (1983).

These values are then used in a multiple linear regression analysis along with the mass concentration values to obtain the scaling factors (Hopke *et al.*, 1980; Severin *et al.*, 1983). This analysis provides an additional test of the quality of the analysis. The regression coefficients should be statistically significant, non-negative, and the rescaled profiles should not sum to

ITEM NAME ID NO A B C D E F G H I J K L M N O P Q R S T U V W X Y
vector 7 7
vector 15 15
vector 2 2
vector 6 6
vector 19 19
vector 20 20
vector 3 3
vector 12 12
vector 11 11
vector 13 13
vector 4 4
vector 17 17
vector 9 9
vector 10 10
vector 5 5
vector 14 14
vector 1 1
vector 16 16
vector 18 18
vector 21 21
vector 8 8
A B C D E F G H I J K L M N O P Q R S T U V W X Y

Figure 17. Dendrogram of the iterated vectors from the TTFA analysis of the Hamilton fine particle data set.

greater than 100%. If there are negative values, either the wrong combination of vectors was chosen or too many factors were retained. If the rescaled vector sums to greater than 100%, then too few vectors have been used. The analysis can then be redone until an appropriate solution has been obtained. The scaling coefficients for the combination given above are presented in Table XIII.

Table XIII. Results of the regression analysis to obtain the scaling factors for the iterative TTFA profiles, Hamilton fine particle data.

No.	Source	Scaling Factor	Uncertainty	t-Ratio
1	vector 4	.2681E+06	1020.	262.8
2	vector 5	.4700E+06	1046.	449.2
3	vector 8	.1818E+06	487.5	373.0
4	vector 16	.1392E+06	854.0	162.9
5	vector 18	.6333E+06	669.0	946.6
6	vector 21	.1005E+06	1042.	96.49

From these results, the rescaled vectors can be examined to determine what sources they represent. These vectors are given in Table XIV. The vectors can be considered in terms of possible sources of airborne particulate matter. Vector 4 has values of Pb and Br as well as Ca. The Br/Pb ratio is 0.30 which is quite typical of aged automotive aerosol. The reaction of the bromine with acidic components of the air such as NH_4HSO_4, H_2SO_4, and HNO_3 leads to the volatilization of bromine as HBr. The lead concentration is 13.2%. This concentration is also quite typical of values observed in the United States before the mandated reductions in lead in gasoline. It appears that there was still substantial use of leaded fuel in Hamilton during this time period.

The second source profile has a high value for chlorine and no other significant value. This source could result from the salting of roads to remove ice and snow. There is also a moderate value for potassium. If there were also non-ferrous metals, this source might be

Table XIV. Source profiles for fine particle Hamilton data (Values are as weight percent).

Element	Motor Vehicles	Salt?	Steel	Soil/ Flyash	Sulfate	Non-Ferrous
Al	2.10	0.97	1.38	11.20	0.65	0.96
As	0.69	0.16	0.00	0.05	0.08	0.00
Br	3.91	0.32	0.52	0.86	0.03	2.69
Ca	3.83	0.00	2.03	4.54	0.00	3.54
Cl	0.03	13.44	0.68	0.47	0.00	6.92
Cr	0.89	0.30	0.03	0.48	0.05	0.08
Cu	1.07	0.22	0.00	0.41	0.08	0.88
Fe	0.52	0.07	39.13	0.01	0.01	3.43
K	2.00	2.06	4.26	3.20	0.16	0.56
Mn	1.00	0.37	2.02	1.20	0.00	1.43
Ni	0.89	0.39	0.00	0.31	0.06	0.02
Pb	13.21	0.81	4.36	2.16	0.12	16.05
P	1.04	0.48	0.27	0.00	0.28	0.00
Rb	0.10	0.36	0.00	0.07	0.12	0.00
Se	0.87	0.18	0.00	0.38	0.03	0.00
Si	2.79	0.18	0.06	44.34	0.02	10.81
Sr	0.35	0.28	0.00	0.14	0.04	0.30
S	0.04	0.03	0.23	0.26	13.89	1.42
Ti	1.01	0.24	0.05	0.64	0.06	0.07
V	0.92	0.29	0.00	0.39	0.05	0.10
Zn	0.05	0.14	0.00	0.73	0.06	50.24

identified as an incinerator. However, both zinc and lead are quite low. Another possible source of chlorine and potassium are wood combustion. It should be remembered that factor analysis identifies elements that covary in time. We normally assume that the covariance is because the elements were emitted by the same source type. The covariance could also arise from other external forces such as local meteorological conditions. Thus, without more detailed understanding of the Hamilton airshed, it is not possible to narrow the range of possible sources of the Cl and K.

The next source has a high iron concentration and very likely represents the nearby steel works. There is high Mn as well as some K, Pb, Ca, and Al associated in the profile. It would require more detailed knowledge of the processes used at the steel works to determine the source(s) of these additional elements. High Al and Si concentrations are found in the "Soil" source profile. The Al/Si ratio appears correct for soil or flyash, but the absolute concentrations are high. However, attempts at a 7 factor solution failed to obtain a satisfactory solution. The low Fe value in this profile is probably due to the presence of the Steel source. Since it is almost a unique factor for Fe, it variation dominates the covariance of iron with the other elements and masks the presence of Fe in the Soil source profile. It is also possible that the "Soil" profile also represents some impact from coal flyash. The volatile elements such as As and Se that are generally enriched in the surface layer of flyash particles are one of the species upon which soil and flyash can be distinguished. Although As is reported above detection limits, the proximity of its K_α emission line (10.532 keV) to the strong L_α line (10.54 keV) means that As is generally not well determined by XRF. In this profile, Se does appear to be enriched. Thus, there is likely to be some contribution of coal flyash to the soil particles in these samples.

The next source is regional sulfate. This source type is universally observed in the fine particle source apportionment. It represents emission sources of SO_2 upwind of Hamilton sufficiently far that the SO_2 has been transformed into $SO_4^=$. The value of 13.2% S in the Sulfate source is somewhat lower than has been observed in other locations where values around 20% S have been observed (Dzubay *et al.*, 1988; Alpert and Hopke, 1981). However, it is similar to values observed at industrialized locations in the St. Louis area (Liu *et al.*, 1982; Severin *et al.*, 1983; Chang *et al.*, 1988). Such a value suggests the presence of carbonaceous secondary aerosol components. Since carbon is not measured by XRF, it cannot be accounted

for directly in the analysis. If it covaries with one or more of the measured elements, the mass contribution from carbon is properly accounted for in the analysis. If there were a carbon source that does not emit any measured elements, then the analysis can encounter considerable difficulties in accounting for all of the measured particle mass (Hopke *et al.*, 1989).

The final profile has high concentrations of Cl, Fe, Pb, Si, and Zn with Pb and Zn representing 66% of the total profile mass. This profile could be from a non-ferrous metal plant. Since the nature of the steel-making processes and the chemical factories in the vicinity of the sampling site are not well known, it is not possible to be more specific in relating this source profile to particular industrial activities.

The mass contributions to the various observed elemental concentrations are presented in Table XV. For example, Pb is approximately equally distributed between the motor vehicles, steel, and non-ferrous sources. It also provides an indication of the quality of fit for each element. The average fit of the elements to the average concentrations are quite good. However, some elements have a large dispersion in the quality of the fit. Aluminum has an

Table XV. Mass contributions of the various sources to the observed elemental concentrations.

Element	Motor Vehicles	Salt?	Steel	Flyash/ Soil	Sulfate	Non-Ferrous	Avg Pred Contrib.	Avg Obs Contrib.	Avg % Error
Al	.176E-01	.267E-01	.410E-01	.138E+00	.122E+00	.679E-02	.352E+00	.349E+00	19.4
As	.575E-02	.442E-02	.000E+00	.621E-03	.145E-01	.000E+00	.253E-01	.191E-01	1608.6
Br	.326E-01	.887E-02	.154E-01	.106E-01	.471E-02	.191E-01	.913E-01	.871E-01	83.4
Ca	.320E-01	.129E-04	.604E-01	.561E-01	.136E-06	.251E-01	.174E+00	.162E+00	41.8
Cl	.263E-03	.371E+00	.201E-01	.581E-02	.121E-04	.490E-01	.446E+00	.447E+00	138.2
Cr	.742E-02	.813E-02	.768E-03	.590E-02	.885E-02	.540E-03	.316E-01	.320E-01	527.0
Cu	.896E-02	.602E-02	.000E+00	.501E-02	.158E-01	.625E-02	.420E-01	.403E-01	254.8
Fe	.433E-02	.184E-02	.116E+01	.672E-04	.172E-02	.243E-01	.120E+01	.120E+01	1.3
K	.167E-01	.567E-01	.127E+00	.395E-01	.290E-01	.395E-02	.273E+00	.261E+00	33.3
Mn	.835E-02	.101E-01	.602E-01	.148E-01	.839E-04	.102E-01	.104E+00	.105E+00	88.6
Ni	.739E-02	.107E-01	.000E+00	.379E-02	.120E-01	.110E-03	.340E-01	.289E-01	1387.8
Pb	.110E+00	.224E-01	.130E+00	.267E-01	.227E-01	.114E+00	.425E+00	.430E+00	5.4
P	.871E-02	.132E-01	.797E-02	.000E+00	.526E-01	.000E+00	.824E-01	.853E-01	106.8
Rb	.828E-03	.983E-02	.000E+00	.904E-03	.231E-01	.000E+00	.346E-01	.187E-01	2068.6
Se	.729E-02	.489E-02	.000E+00	.470E-02	.585E-02	.000E+00	.227E-01	.210E-01	2901.8
Si	.233E-01	.490E-02	.190E-02	.547E+00	.457E-02	.766E-01	.658E+00	.661E+00	4.4
Sr	.292E-02	.781E-02	.000E+00	.176E-02	.800E-02	.210E-02	.226E-01	.144E-01	1818.5
S	.338E-03	.833E-03	.671E-02	.326E-02	.260E+01	.101E-01	.262E+01	.262E+01	.1
Ti	.840E-02	.675E-02	.136E-02	.785E-02	.109E-01	.502E-03	.357E-01	.362E-01	303.1
V	.766E-02	.796E-02	.000E+00	.478E-02	.923E-02	.695E-03	.303E-01	.271E-01	957.7
Zn	.381E-03	.378E-02	.684E-04	.896E-02	.107E-01	.356E+00	.380E+00	.379E+00	5.3

average point-by-point uncertainty of 19.4% whereas arsenic has a value of over 1600%. Thus, there is much poorer ability to predict an individual value of arsenic than an individual value of aluminum. In general the elements with large uncertainties are those low in concentration and therefore, have many values near the detection limits. The results presented here are fairly typical of those observed in the kind of analysis.

Finally the contribution of the source types to the airborne particle mass contributions can be determined. The mass contributions in $\mu g/m^3$ are plotted as a function of sampling date in Figures 18-23. There is not easily discerned pattern in these results. However, the intermittent nature of the sampling schedule, particularly in the latter part of the time period, makes it difficult to have a sufficient time record to discern temporal patterns in the results.

Thus, the TTFA has provided an indication of the types of sources contributing particles to the observed airborne concentrations and the amount of mass contributed by those sources and represent the type of results that can be expected of such an analysis. At this time, the results for the coarse particle samples are still being analyzed. We have not yet found a suitable solution. Once an acceptable solution has been obtained for the coarse samples, we will be able to examine the overall pattern of source contributions to the observed airborne particulate concentrations in downtown Hamilton.

Summary

The objective of this initial grant was to begin to develop methods to utilize the rich resource of the Ministry of the Environment's substantial data base of environmental measurements so that a better understanding of the chemical and physical processes that give rise to those observations could be obtained. We have made substantial progress in developing three-mode factor analysis into a promising tool for environmental data analysis. It has been shown that existing methods of combining compositional data with meteorological information in the form of air parcel back trajectories could be applied to precipitation data. The use of target transformation factor analysis to identify and apportion sources of urban scale airborne particulate matter has also been explored. We have thus made substantial progress during the two years of this grant that has resulted in two journal publications being published (Zeng and Hopke, 1989a; Zeng and Hopke, 1990). During the next three years, we plan to complete the

Hamilton study and to continue to explore new methods or innovative uses of existing analytical tools to provide more information from the data already collected by the Ministry.

Acknowledgement

We would like to thank Dr. Walter Chan who was our project manager on this grant. His interest in and support of our studies is greatly appreciated. We would also like to thank those individuals in the Air Resources Branch including Neville Reid, Peter Steer, Al Tang, and Diane Green who help provide the data and other information necessary to perform our studies and for a number of interesting conversations regarding the work, directions being taken, and the interpretation of the results. We look forward to continued collaboration in the future.

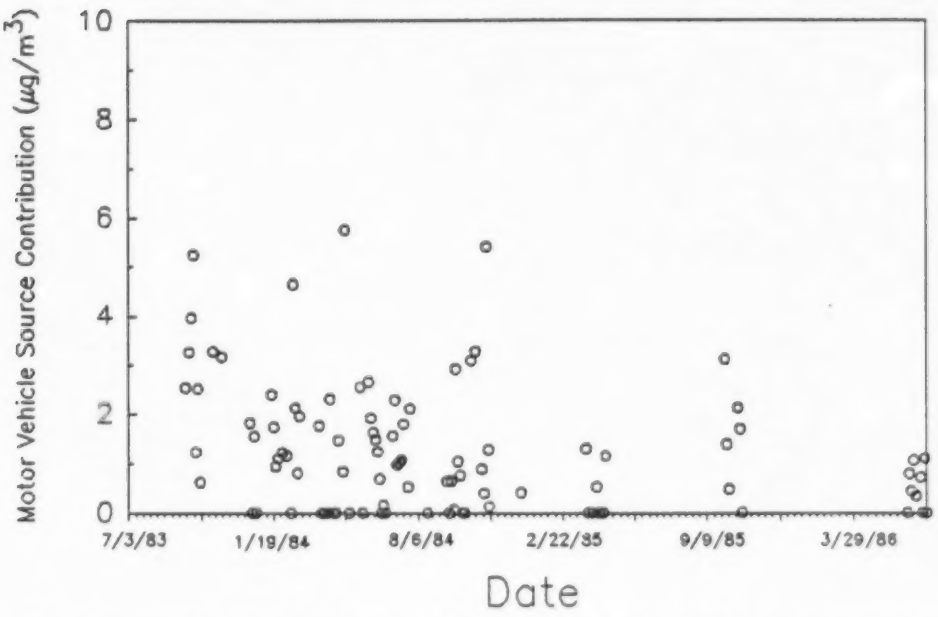

Figure 18. Mass contribution for motor vehicles to the fine particle mass concentration in Hamilton.

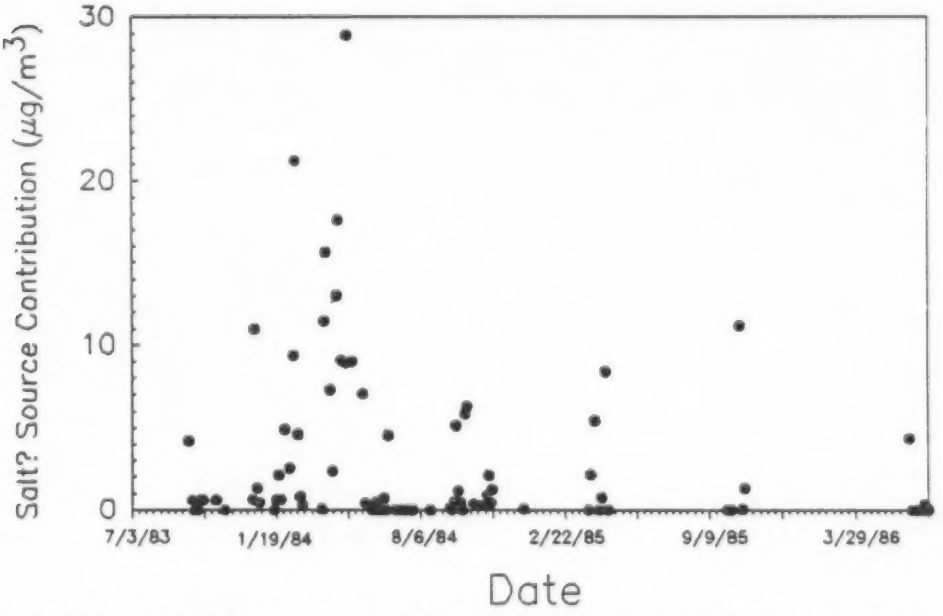

Figure 19. Mass contribution for the Salt? source to the fine particle mass concentration in Hamilton.

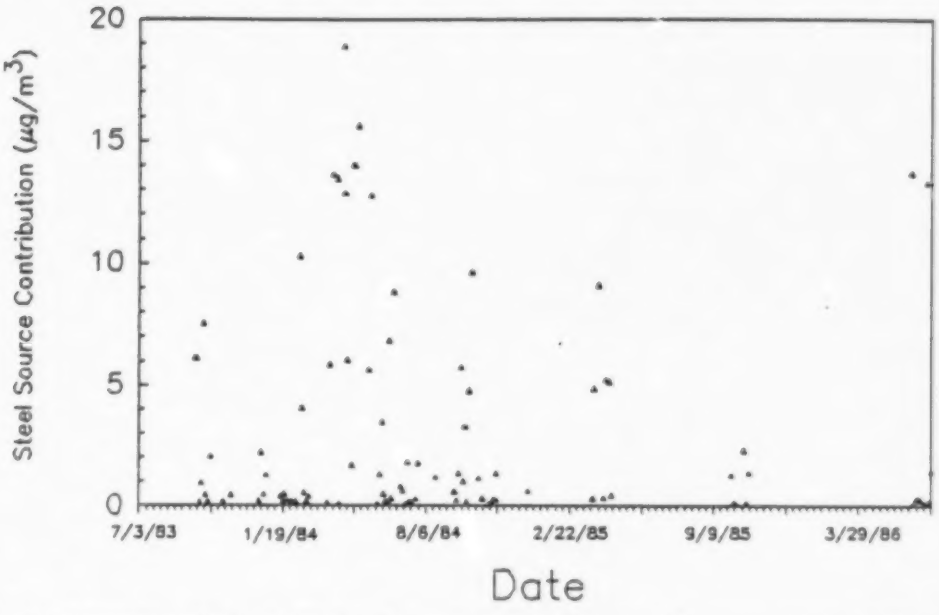

Figure 20. Mass contribution for steel source to the fine particle mass concentration in Hamilton.

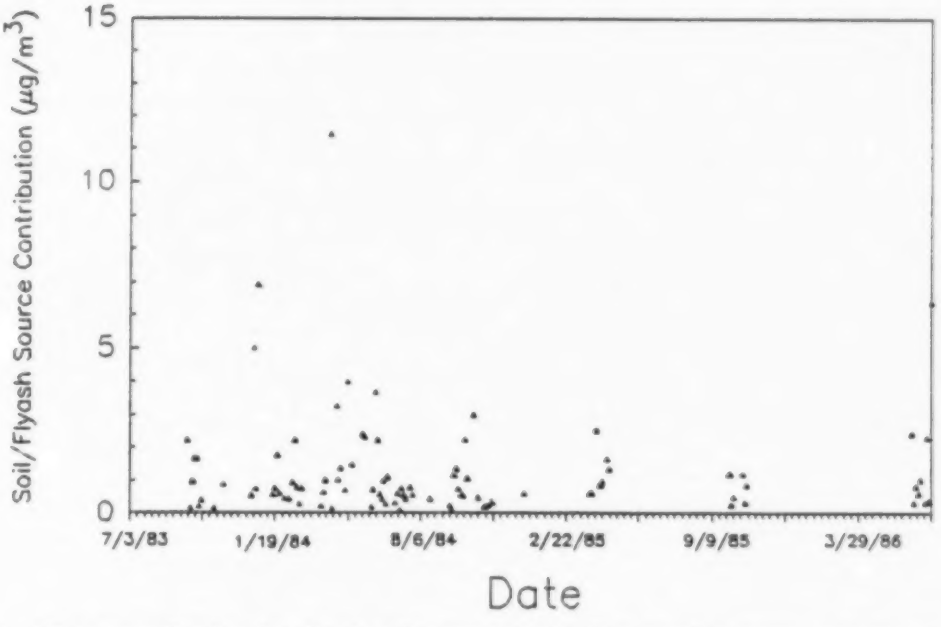

Figure 21. Mass contribution for flyash/soil source to the fine particle mass concentration in Hamilton.

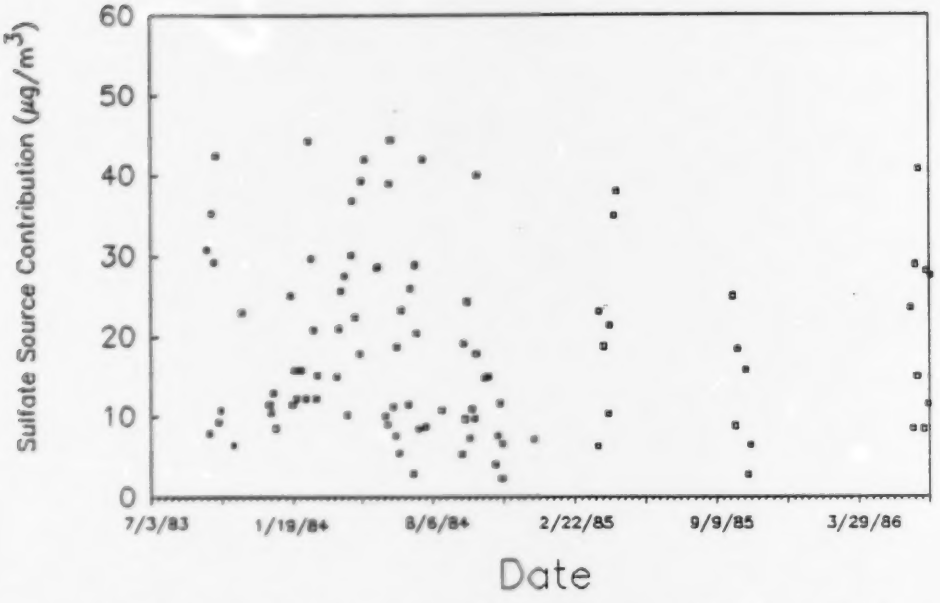

Figure 22. Mass contribution of the regional sulfate aerosol to the fine particle mass concentration in Hamilton.

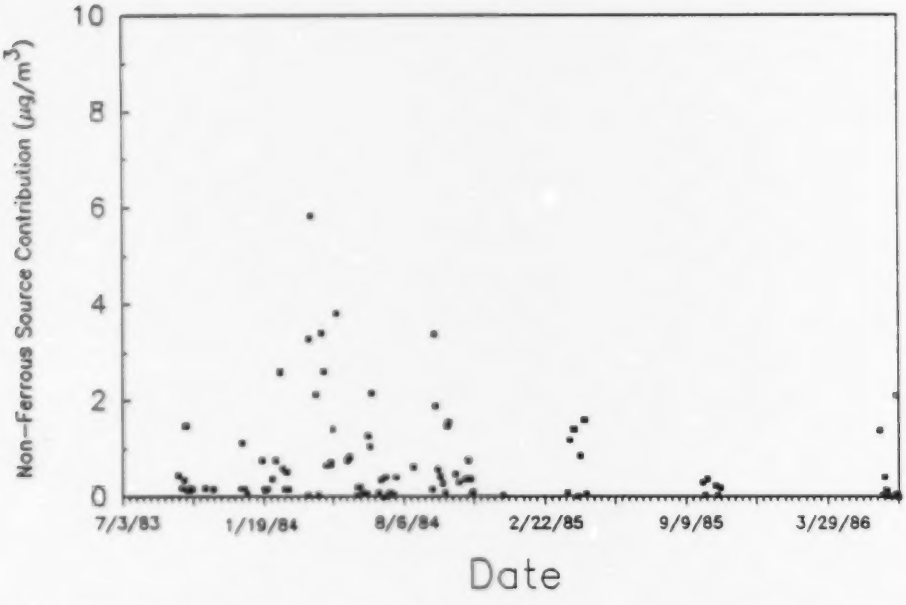

Figure 23. Mass contribution for the non-ferrous metal source to the fine particle mass concentration in Hamilton.

References

Alpert, D.J. and P.K. Hopke (1981) A Determination of the Sources of Airborne Particles Collected During the Regional Air Pollution Study, *Atmospheric Environ.* 15:675-687.

Blackith, R.E. and R.A. Reyment (1971) *Multivariate Morphometrics*, Academic Press, London.

Bresch, J. E., Ashbaugh, L. L., Henmi, T., and Reiter, E. R. (1984), Comparison of a Single-Layer and a Multilayer Transport Model for Residence Time Analysis, 77th Annual Air Pollution Control Assoc. Meeting, San Francisco, California.

Chan, W. H., Orr, D. B., Bardswick, W. S., and Vet, R. J.(1985), Acidic Precipitation in Ontario Study (APIOS); An Overview: The Event Wet/Dry Deposition Network (1st revised edition), Ontario Ministry of the Environment, Report # ARB-142-85-AQM, APIOS-025-85.

Chan, W.H. (1988) private communication.

Chang, S.N., P.K. Hopke, G.E. Gordon, and S.W. Rheingrover (1988) Target Transformation Factor Analysis of Airborne Particulate Samples Selected by Wind-Trajectory Analysis, *Aerosol Sci. Technol.* 8:63-80.

Dzubay, T.G., R.K. Stevens, and L.W. Richards (1979) Composition of Aerosols over Los Angeles Freeways, *Atmospheric Environ.*, 13:653-659.

Dzubay, T.G., R.K. Stevens, W.D. Balfour, H.J. Williamson, J.A. Cooper, J.E. Core, R.T. DeCesar, E.R. Crutcher, S.L. Dattner, B.L. David, S.L. Heisler, J.J. Shah, P.K. Hopke, and D.L. Johnson (1984) Interlaboratory Comparison of Receptor Model Results for Houston Aerosol *Atmospheric Environ.* 18:1555-1566.

Dzubay, T.G., R.K. Stevens, G.E. Gordon, I. Olmez, A.E. Sheffeld, and W.J. Courtney (1988) A Composite Receptor Method Applied to Philadelphia Aerosol, *Environ. Sci. Technol.* 22:46-52.

EPA (1986) PM10 SIP Development Guideline, Report No. EPA450/2-86-001, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Harman, H.H. (1976) *Modern Factor Analysis*, 3rd Edition, University of Chicago Press, Chicago.

Harrison, R.M. and W.T. Sturges (1983) The Measurement and Interpretation of Br/Pb Ratios in Airborne Particles, *Atmospheric Environ.*, 17:311-328.

Henry, R.C., C.W. Lewis, P.K. Hopke, and H.J. Williamson (1984) Review of Receptor Model Fundamentals, *Atmospheric Environ.* 18:1507-1515.

Hopke, P.K. (1985) *Receptor Modeling in Environmental Chemistry*, J. Wiley and Sons, New York.

Hopke, P.K. (1988) Target Transformation Factor Analysis as an Aerosol Mass Apportionment Method: A Review and Sensitivity Analysis, *Atmospheric Environ.* 22:1777-1792.

Hopke, P.K. (1989) Target Transformation Factor Analysis, P.K. Hopke, *Chemometrics and Intelligent Laboratory Systems* 6:7-19.

Hopke, P.K. and S. Dharmavaram, (1986) Recent Improvements to FANTASIA: A Target Transformation Factor Analysis Program, *Computers & Chemistry* 10:163-164.

Hopke, P.K., R.E. Lamb and D.F.S. Natusch (1980) Multielemental Characterization of Urban Roadway Dust, *Environ. Sci. Technol.* 14:164-172.

Hopke, P.K., D.J. Alpert, and B.A. Roscoe (1983) FANTASIA - A Program for Target Transformation Factor Analysis to Apportion Sources in Environmental Samples, *Computers and Chemistry* 7:149-155.

Hopke, P.K., W. Wlaschin, S. Landsberger, C. Sweet, and S.J. Vermette (1988) The Source Apportionment of PM10 in South Chicago, in *PM-10: Implementation of Standards*, C.V. Mathai and D.H. Stonefield, Eds., Air Pollution Control Association, Pittsburgh, PA, pp. 484-494.

Joreskog, K.G., J.E. Klovan, and R.A. Reyment (1976) *Geological Factor Analysis*, Elsevier Scientific Publishing Co., Amsterdam.

Kaiser, H.F. (1958) The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, 23:187.

Keeler, G. J. and P.J. Samson (1987) Testing Source-receptor Relationships for Trace Elements, Presented at the 80th Annual Air Pollution Control Association Meeting, New York, NY, July 1987.

Kroonenberg, P.M. (1983) *Three-Mode Principal Component Analysis*, DSWO Press, Leiden, The Netherlands.

Kroonenberg, P.M. and P. Brouwer (1985) User's Guide to TUCKALS3, Version 4.0, Department of Education, University of Leiden, The Netherlands.

Liu, C.K., B.A. Roscoe, K.G. Severin, and P.K. Hopke (1982) The Application of Factor Analysis to Source Apportionment of Aerosol Mass, *Am. Ind. Hyg. Assoc. J.* 43:314-318.

Lowenthal, D. H., K.R. Wunschel, and K.A. Rahn (1988) Tests of Regional Element Tracers of Pollution Aerosols. 1. Distinctness of Regional Signatures, Stability during Transport, and Empirical Validation, *Environ. Sci. Technol.* 22:413-420.

Lowenthal, D. H. and K. A. Rahn (1988), Tests of Regional Element Tracers of Pollution Aerosols. 2. Sensitivity of Signatures and Apportionments to Variations in Operating Parameters, *Environ. Sci. Technol.* 22:420-426.

Malinowski, E.R. and D.G. Howery (1980) *Factor Analysis in Chemistry*, J. Wiley & Sons Inc., New York.

Malm, W.C., C.E. Johnson, and J.F. Bresch (1986) Application of Principal Component Analysis for Purposes of Identifying Source-Receptor Relationships. In *Receptor Methods for Source Apportionment*, T.G. Pace, Ed., Air Pollution Control Association, Pittsburgh, PA. pp. 127-148.

Massart, D.L. and L. Kaufman (1983) *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, J. Wiley & Sons, Inc., New York.

Miller, M.S., S.K. Friedlander, and G.M. Hidy (1972) A Chemical Element Balance for the Pasadena Aerosol, *J. Colloid Interface Sci.*, 39:65-176.

Olson, M. P., Oikawa, K. K., and Macafee, A. W. (1978) A Trajectory Model applied to the Long-Range Transport of Air Pollutants, Report of Atmos. Environ. Service, Downsview, Ontario.

Rahn, K. A. and Lowenthal, D. H. (1984), Elemental Tracers of Distant Regional Pollution Aerosols, *Science*, 223:132-139.

Rahn, K. A. and Lowenthal, D. H. (1985), Pollution Aerosol in the Northeast: Northeastern-Midwestern Contribution, *Science*, 228:275-284.

Rheingrover, S.W. and G.E. Gordon (1988) Wind-Trajectory Method for Determining Compositions of Particles from Major Air Pollution Sources, *Aerosol Sci. Technol.* 8:29-61.

Roscoe, B.A. and P.K. Hopke (1981) Comparison of Weighted and Unweighted Target Transformation Rotations in Factor Analysis, *Computers & Chem.*, 5:1-7.

Roscoe, B.A., P.K. Hopke, S.L. Dattner and J.M. Jenks (1982) The Use of Principal Component Factor Analysis to Interpret Particulate Compositional Data Sets, *J. Air Pollution Control Assoc.* 32:637-642.

Severin, K.G., B.A. Roscoe, and P.K. Hopke (1983) The Use of Factor Analysis in Source Determination of Particulate Emissions, *Particulate Science and Technology*, 1:183-192.

Stevens, R.K. and T.G. Pace (1984) Overview of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II), *Atmospheric Environ.*, 18:1499-1506.

Winchester, J.W. and G.D. Nifong (1971) Water Pollution in Lake Michigan by Trace Elements from Pollution Aerosol Fallout, *Water, Air, Soil Pollut.*, 1:50-64.

Zeng, Y. and P.K. Hopke (1988) Approaches to Study the Sources of Acid Precipitation in Ontario, Canada: A Technical Description, Report to The Ontario Ministry of the Environment, Institute for Environmental Studies, University of Illinois at Urbana-Champaign, September 1988.

Zeng, Y. and P.K. Hopke (1989a) Three-Mode Factor Analysis: A New Multivariate Method for Analyzing Spatial and Temporal Composition Variation. In *Receptor Models in Air Resources Management*, Air Pollution Control Association, Pittsburgh, PA, pp. 173-179.

Zeng, Y. and P.K. Hopke (1989b) A Study on the Sources of Acid Precipitation in Ontario, Canada, *Atmospheric Environ.* 23:1499-1509.

Zeng, Y. and P.K. Hopke (1990) Methodological Study Applying Three-Mode Factor Analysis to Three-Way Chemical Data Sets, *Chemometrics and Intelligent Laboratory Systems* (in press).